

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 01:11:34 ; Search time 181.2 Seconds
(without alignments)
194.250 Million cell updates/sec

Title: US-09-462-480-5
Perfect score: 492
Sequence: 1 MAEMKTDAAATLGSEAGNFER.....VOYSRADEEQOALSSOMGF 100

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	492	100.0	100	15	US-09-116-492-5
2	492	100.0	100	18	Sequence 5, Appli
3	486	98.8	100	15	US-09-462-480-5
4	486	98.8	100	1	PCT-US99-03265-110
5	486	98.8	100	1	PCT-US99-03268-115
6	486	98.8	100	11	US-08-729-622-110
7	486	98.8	100	11	US-08-730-510-115
			100	13	US-08-942-341-110

8	486	98.8	100	13	US-08-942-578-115
9	486	98.8	100	14	US-09-024-753-110
10	486	98.8	100	14	US-09-025-197-115
11	486	98.8	100	14	US-09-072-596-110
12	486	98.8	100	14	US-09-072-967-115
13	486	98.8	100	20	US-09-615-947-1
14	486	98.8	100	21	US-09-724-685-115
15	486	98.8	100	24	US-10-084-843-115
16	481	97.8	802	1	PCT-US99-03265-209
17	481	97.8	802	1	PCT-US99-03265-346
18	481	97.8	802	1	PCT-US99-03268-214
19	481	97.8	802	1	PCT-US99-03268-351
20	481	97.8	802	13	US-08-942-341-209
21	481	97.8	802	13	US-08-942-578-214
22	481	97.8	802	14	US-09-024-753-209
23	481	97.8	802	14	US-09-025-197-214
24	481	97.8	802	14	US-09-072-596-209
25	481	97.8	802	14	US-09-072-596-346
26	481	97.8	802	14	US-09-072-967-214
27	481	97.8	802	14	US-09-072-967-351
28	481	97.8	802	16	US-09-287-849-10
29	481	97.8	802	24	US-10-084-843-214
30	481	97.8	802	24	US-10-084-843-351
31	462	93.9	95	1	PCT-US99-03265-88
32	462	93.9	95	1	PCT-US99-03268-88
33	462	93.9	95	10	US-08-658-800-89
34	462	93.9	95	10	US-08-659-683-88
35	462	93.9	95	10	US-08-680-573-89
36	462	93.9	95	10	US-08-680-574-88
37	462	93.9	95	11	US-08-729-622-89
38	462	93.9	95	11	US-08-730-510-88
39	462	93.9	95	13	US-08-942-341-89
40	462	93.9	95	13	US-08-942-578-88
41	462	93.9	95	14	US-09-024-753-89
42	462	93.9	95	14	US-09-025-197-88
43	462	93.9	95	14	US-09-072-596-89
44	462	93.9	95	14	US-09-072-967-88
45	462	93.9	95	21	US-09-724-685-88

ALIGNMENTS

RESULT 1

US-09-116-492-5
; Sequence 5, Application US/09116492
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN
; TITLE OF INVENTION: FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY
; TITLE OF INVENTION: ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS METHODS USING
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492
; CURRENT FILING DATE: 1998-07-16
; EARLIER APPLICATION NUMBER: 60/052,631
; EARLIER FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-116-492-5

Query Match 100.0%; Score 492; DB 15; Length 100;
Best Local Similarity 100.0%; Pred. No. 3.8e-44;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOEAGNFERISGLDKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60
Db 1 MAEMKTDATLGOEAGNFERISGLDKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60
QY 61 AANKKOELDEISTNIRAGVQVSRADEEQOQALSSQMGF 100
Db 61 AANKKOELDEISTNIRAGVQVSRADEEQOQALSSQMGF 100

RESULT

US-09-462-480-5
; Sequence 5, Application US/09462480
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER BIRK
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYC
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS W
; FILE REFERENCE: 0660-0165-0XPCT
; CURRENT APPLICATION NUMBER: US/09/462,480
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: PCT/TB98/01091
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/052,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-462-480-5

Query Match 100.0%; Score 492; DB 18; Length 100;
Best Local Similarity 100.0%; Pred. No. 3.8e-44;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOEAGNFERISGLDKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60
Db 1 MAEMKTDATLGOEAGNFERISGLDKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60
QY 61 AANKKOELDEISTNIRAGVQVSRADEEQOQALSSQMGF 100
Db 61 AANKKOELDEISTNIRAGVQVSRADEEQOQALSSQMGF 100

RESULT

PCT-US99-03265-110
; Sequence 110, Application PC/TUS9903265
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US99/03265
; FILING DATE: 17-FEB-1999
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/024,753
; FILING DATE: 18-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9532-0023-228
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US99-03265-110

Query Match 98.8%; Score 486; DB 1; Length 100;
Best Local Similarity 99.0%; Pred. No. 1.7e-43;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MAEMKTDATLGOEAGNFERISGLDKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60
Db 1 MAEMKTDATLGOEAGNFERISGLDKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60
QY 61 AANKKOELDEISTNIRAGVQVSRADEEQOQALSSQMGF 100
Db 61 AANKKOELDEISTNIRAGVQVSRADEEQOQALSSQMGF 100

RESULT

PCT-US99-03268-115
; Sequence 115, Application PC/TUS9903268
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US99/03268
; FILING DATE: 17-FEB-1999
; CLASSIFICATION: 406
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/025,197
; FILING DATE: 18-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9532-0013-228
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:

TOPOLOGY: linear
PCT-US99-03268-115

Query Match 98.8%; Score 486; DB 1; Length 100;
Best Local Similarity 99.0%; Pred. No. 1.7e-43;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWKGAAGTAAQAAVVRQF 60
|||||

Db 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWKGAAGTAAQAAVVRQF 60
|||||

QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
|||||

Db 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
|||||

RESULT 5

US-08-729-622-110
; Sequence 110, Application US/08729622
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas H.
; APPLICANT: Twardzik, David R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 132
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/729,622
; FILING DATE: 11-OCT-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-729-622-110

Query Match 98.8%; Score 486; DB 1; Length 100;
Best Local Similarity 99.0%; Pred. No. 1.7e-43;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWKGAAGTAAQAAVVRQF 60
|||||

Db 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWKGAAGTAAQAAVVRQF 60
|||||

QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
|||||

Db 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

RESULT 6

US-08-730-510-115
; Sequence 115, Application US/08730510
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 137
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/730,510
; FILING DATE: 27-AUG-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-730-510-115

Query Match 98.8%; Score 486; DB 1; Length 100;
Best Local Similarity 99.0%; Pred. No. 1.7e-43;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWKGAAGTAAQAAVVRQF 60
|||||

Db 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWKGAAGTAAQAAVVRQF 60
|||||

QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
|||||

Db 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
|||||

RESULT 7

US-08-942-341-110
; Sequence 110, Application US/08942341
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.

APPLICANT: Lodes, Michael J.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,341
FILING DATE: 01-OCT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C7
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 110:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
US-08-942-341-110

Query Match 98.8%; Score 486; DB 13; Length 100;
Best Local Similarity 99.0%; Pred. No. 1.7e-43;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTDQIDVESTAGSLQGWGAAAGTAAQAAYVRFQE 60
|||||
Db 1 MAEMKTDAAATLGOEAGNFERISGDLKTDQIDVESTAGSLQGWGAAAGTAAQAAYVRFQE 60
|||||

QY 61 AANKQKQELDEISTNIRAGVQYSRADEEQQALSSOMGF 100
|||||
Db 61 AANKQKQELDEISTNIRAGVQYSRADEEQQALSSOMGF 100
|||||

RESULT 8
US-08-942-578-115
Sequence 115, Application US/08942578
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 214
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,578
FILING DATE: 01-OCT-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C7
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 115:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
US-08-942-578-115

Query Match 98.8%; Score 486; DB 13; Length 100;
Best Local Similarity 99.0%; Pred. No. 1.7e-43;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTDQIDVESTAGSLQGWGAAAGTAAQAAYVRFQE 60
|||||
Db 1 MAEMKTDAAATLGOEAGNFERISGDLKTDQIDVESTAGSLQGWGAAAGTAAQAAYVRFQE 60
|||||

QY 61 AANKQKQELDEISTNIRAGVQYSRADEEQQALSSOMGF 100
|||||
Db 61 AANKQKQELDEISTNIRAGVQYSRADEEQQALSSOMGF 100
|||||

RESULT 9
US-09-024-753-110
Sequence 110, Application US/09024753
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 236
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/024,753
FILING DATE: 18-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C8
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 110:

1 MAEUKTDAATGGEAGNFEERTSSDUKTOIDOVESTAGSLOGOWRGAAGTAAOAAVVREOF 60

;
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G

APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedzick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 115:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
US-09-072-967-115

Query Match 98.8%; Score 486; DB 14; Length 100;
Best Local Similarity 99.0%; Pred. No. 1.7e-43;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOE 60
DB 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOE 60
QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGE 100
DB 61 AANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGE 100
RESULT 13
US-09-615-947-1
Sequence 1, Application US/09615947
GENERAL INFORMATION:
APPLICANT: ANDERSEN, Peter et al.
TITLE OF INVENTION: Tuberculosis vaccine and diagnostics
TITLE OF INVENTION: based on the Mycobacterium tuberculosis esat-6 gene family
FILE REFERENCE: 0459-0462P
CURRENT APPLICATION NUMBER: US/09/615,947
CURRENT FILING DATE: 2001-06-04
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 100
TYPE: PRT
ORGANISM: M.Tuberculosis
US-09-615-947-1

Query Match 98.8%; Score 486; DB 20; Length 100;
Best Local Similarity 99.0%; Pred. No. 1.7e-43;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOE 60
DB 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOE 60
QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGE 100
DB 61 AANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGE 100
RESULT 14
US-09-724-685-115
Sequence 115, Application US/09724685
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
TITLE OF INVENTION: Compounds and Methods for
Immunotherapy and Diagnosis of Tuberculosis
NUMBER OF SEQUENCES: 155
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/724,685
FILING DATE: 28-Nov-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/523,436
FILING DATE: 01-SEP-1995
APPLICATION NUMBER: US 08/533,634
FILING DATE: 22-SEP-1995
APPLICATION NUMBER: US 08/620,874
FILING DATE: 22-MAR-1996
APPLICATION NUMBER: US 08/659,683
FILING DATE: 05-JUN-1996
APPLICATION NUMBER: US 08/680,574
FILING DATE: 12-JUL-1996
APPLICATION NUMBER: WO PCT/US96/14674
FILING DATE: 30-AUG-1996
APPLICATION NUMBER: US 08/730,511
FILING DATE: 11-OCT-1996
APPLICATION NUMBER: US 08/818,112
FILING DATE: 13-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 014058-008561US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 115:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 115:

US-09-724-685-115

Query Match 98.8%; Score 486; DB 21; Length 100;
Best Local Similarity 99.0%; Pred. No. 1.7e-43;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60
Db 1 MAEMKTDATLQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60
QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
Db 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

RESULT 15

US-10-084-843-115
; Sequence 115, Application US/10084843
; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 115:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 115:

US-10-084-843-115

Query Match 98.8%; Score 486; DB 24; Length 100;
Best Local Similarity 99.0%; Pred. No. 1.7e-43;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60

Db 1 MAEMKTDATLQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60
QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
Db 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

Search completed: July 22, 2002, 01:16:48
Job time: 314 sec

/SIDS1/gcgdata/hold-geneseg/genseq-emb1/AA1988.DAT.AA#81697	+ 118.00	227.93	0.0
/SIDS1/gcgdata/hold-geneseg/genseq-emb1/AA1999.DAT.AA#39125	+ 118.00	227.93	0.0
/SIDS1/gcgdata/hold-geneseg/genseq-emb1/AA1999.DAT.AA#39127	+ 118.00	227.93	0.0
/SIDS1/gcgdata/hold-geneseg/genseq-emb1/AA1997.DAT.AA#32455	+ 114.00	220.22	0.0
/SIDS1/gcgdata/hold-geneseg/genseq-emb1/AA1998.DAT.AA#81693	+ 114.00	220.22	0.0

```

seq_name: /SIBS179Cguata/nord_geneseq/geneseq_emb1/AA19393.DAT:AA103705
seq_documentation_block:
ID   AAY03705 standard; Protein; 100 AA.
XX
XX
AC   AAY03705;
XX
XX   07-JUN-1999 (first entry)
XX
XX   M. tuberculosis LHP polypeptide.
DE
XX
XX   ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;
KW   immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;
KW   immune response.

```

xx
xx OS Mycobacterium tuberculosis.
xx WO9904005-A1.
xx PN
xx XX 28-JAN-1999.
xx PD
xx PF 16-XIII-1998: q8WO-TB01091.

XX (INSP) INST PASTEUR.
 PA (STAT-) STATENS SERUM INST.
 XX
 XX Andersen P, Berthet F, Gicquel B, Rasmussen PB;
 PI
 XX WPI; 1999-132249/11.
 XX N-PSDB; AAX29168, AAX29171.
 DR
 XX
 XX New nucleic acid containing regulator and LHP gene of Mycobacterium
 PT tuberculosis - useful in vaccines, for diagnosis, and for expression
 PT of heterologous proteins
 PT
 XX
 XX Claim 17; Page 64; 88pp; English.
 PS
 XX
 CC The present invention is directed to a polynucleotide carrying the
 CC regulatory expression signals of the ESAT-6 protein as well as an open
 CC reading frame coding for an antigenic protein LHP from Mycobacterium
 CC tuberculosis. Host cells comprising the polynucleotide are used for the
 CC recombinant expression of the protein. The recombinant polypeptide can be
 CC used as immunogens and vaccines, to protect against bacteria of the
 CC M. tuberculosis complex in humans or animals (the vaccines may include
 CC other immunogenic proteins of the bacteria or their fragments,
 CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by
 CC detection of specific antibodies. The regulatory region present in the
 CC polynucleotide may be used to express almost any heterologous protein in
 CC mycobacteria, particularly as a fusion with polyhistidine. The two
 CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to
 CC provide a synergistic increase in ability to induce a protective immune
 CC response. The present sequence represents the LHP polypeptide.
 CC
 XX SQ Sequence 100 AA;

 alignment_scores: Length: 100
 Quality: 492.00
 Ratio: 4.920
 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

 alignment_block:
 US-09-462-480-4 x AAY03705 ..

 Align seq 1/1 to: AAY03705 from: 1 to: 100

```
alignment_scores:
  Quality: 492.00      Length: 100
  Ratio: 4.920        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-462-480-4 x AAY03705      .
  Align seq 1/1 to: AAY03705      from: 1 to: 100
```

```
1 ATGGCAGAGATGAAGACCGATCCCGTACCCCTCGGCGAGGAGGTA 50
|||||
1 MetAlaGluMetLysThrAspAlaIaThrLeuGlyGlnGluAlaGlyAs 17
|||||
51 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCAAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
|||||
101 CGACGGCAGGTTCTGTCAGGCGCAGTGGCGGCGCGGCGGAGCGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
|||||
151 GCCCAGCGCGCGTGTGCGCTTCCAAGACGAGCAATTAAGCAGAAGCA 200
|||||
51 AlaGlnAlaAlaValIaValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
|||||
201 GGAATCGACGAGATCTCGACGAATATTCTGTCAGCGCGCTCCCAAT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
|||||
251 CGAGGGCGCAGCAGCAGCAGCAGCGCTGTCTCGCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
|||||
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81706

seq_documentation_block:

ID: AAW81706 standard; Protein; 100 AA.

XX AAW81706;

XX AC

XX DT

XX 27-JAN-1999 (first entry)

DE M. tuberculosis immunogenic polypeptide Tb38-IN.

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;

KW vaccine; pharmaceutical; infection; diagnosis.

KW Mycobacterium tuberculosis.

OS

XX WO9816646-A2.

XX PD

XX 23-APR-1998.

XX PF

XX 07-OCT-1997; 97WO-US18293.

XX PR

XX 13-MAR-1997; 97US-0818112.

XX PR

XX 11-OCT-1996; 96US-0730510.

XX PA

XX (CORI-) CORIXA CORP.

XX PI

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-261042/23.

XX DR

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used

XX to develop products for the detection of M. tuberculosis infection

XX and for diagnosis, treatment and prevention of tuberculosis

XX Example 3B; Page 138-139; 230pp; English.

XX PS

XX This sequence represents an immunogenic portion of a soluble

XX Mycobacterium tuberculosis (MT) antigen which can be used in a method

CC for inducing protective immunity against tuberculosis (TB). This

CC sequence can be formulated into vaccines and/or pharmaceutical

CC compositions for immunising against M. tuberculosis infection or may

CC be used for the diagnosis of tuberculosis.

XX CC

XX Sequence 100 AA;

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:

US-09-462-480-4 x AAW81706

Align seg 1/1 to: AAW81706 from: 1 to: 100

```
1 ATGGCAGAGATGAAGACCGATCCCGTACCCCTCGGCGAGGAGGTA 50
|||||
1 MetAlaGluMetLysThrAspAlaIaThrLeuAlaGlnGluAlaGlyAs 17
|||||
51 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCAAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
|||||
101 CGACGGCAGGTTCTGTCAGGCGCAGTGGCGGCGCGGCGGAGCGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
|||||
151 GCCCAGCGCGCGTGTGCGCTTCCAAGACGAGCAATTAAGCAGAAGCA 200
|||||
51 AlaGlnAlaAlaValIaValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
|||||
201 GGAATCGACGAGATCTCGACGAATATTCTGTCAGCGCGCTCCCAAT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
|||||
251 CGAGGGCGCAGCAGCAGCAGCAGCGCTGTCTCGCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
|||||
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64339

seq_documentation_block:

ID: AAW64339 standard; Protein; 100 AA.

XX AAW64339;

XX AC

XX DT

XX 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen Tb38-IN.

DE Tuberculosis; infection; diagnosis; antigen; Tb37-FL.

KW Mycobacterium tuberculosis strain H37Rv.

XX OS

XX WO9816645-A2.

XX PN

XX 23-APR-1998.

XX PF

XX 07-OCT-1997; 97WO-US18214.

XX PR

XX 13-MAR-1997; 97US-0818111.

XX PR

XX 11-OCT-1996; 96US-0729622.

XX PA

XX (CORI-) CORIXA CORP.

XX PI

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-251292/22.

XX DR

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used

PT to develop products for the detection of M. tuberculosis infection

PT and diagnosis of tuberculosis

XX Example 3; Page 145; 250pp; English.

XX PS

XX This polypeptide comprises a partial sequence of Mycobacterium

CC tuberculosis antigen Tb38-IN. It is encoded by genomic DNA isolated

CC from a M. tuberculosis strain H37Rv genomic library using a probe
CC derived from clone Tb38-1 (see AAV44384). The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of
CC a soluble M. tuberculosis antigen, or an immunogenic portion of an
CC M. tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using
CC these polypeptides, antibodies or oligonucleotide probes and
CC primers, for the diagnosis of tuberculosis.
XX
SQ Sequence 100 AA;

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000
alignment_block:
US-09-462-480-4 x AAW64339 ..

Align seg 1/1 to: AAW64339 from: 1 to: 100

1 ATGCGAGAGATGAAGACCGATCGCGCTACCTCGGCGAGGAGCGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGCGCACCTGAAACCCAGATCGACAGGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGAGCGGAGGTTCGTCAGGCGCAGTCGCGCGCGCGCGCGCGCGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGGCGCGGTGGTGGCTTCCAAAGACGACGACCAATAGCAGAAGCA 200
|||||
51 AlaGlnAlaAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAACCTGACGAGATCTCGACGAATATTCGTCAGCGCGCGCTCCAATACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
251 CGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
|||||
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV39136
seq_documentation_block:
ID AAY39136 standard; Protein; 100 AA.
XX
AC AAY39136;

XX 05-NOV-1999 (first entry)

DE M. tuberculosis antigen Tb38-IN amino acid sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW Immune response; skin test.

XX Mycobacterium tuberculosis.

XX W09942076-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99NO-US03268.

XX 05-MAY-1998; 98US-0072967.

PR 18-FEB-1998; 98US-0025197.
XX (CORI-) CORIXA CORP.
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MU, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1999-527409/44.
XX New antigens from Mycobacterium tuberculosis useful in diagnostic
XX skin tests and protective or therapeutic vaccines or compositions
XX
XX Example 3; Page 133-134; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.

XX Sequence 100 AA;

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x AAY39136 ..

Align seg 1/1 to: AAY39136 from: 1 to: 100

1 ATGCGAGAGATGAAGACCGATCGCGCTACCTCGGCGAGGAGCGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGCGCACCTGAAACCCAGATCGACAGGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGAGCGGAGGTTCGTCAGGCGCAGTCGCGCGCGCGCGCGCGCGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGGCGCGGTGGTGGCTTCCAAAGACGACGACCAATAGCAGAAGCA 200
|||||
51 AlaGlnAlaAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAACCTGACGAGATCTCGACGAATATTCGTCAGGCGCGCTCCAATACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
251 CGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
|||||
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV38993
seq_documentation_block:
ID AAY38993 standard; Protein; 100 AA.
XX
AC AAY38993;

XX 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen protein Tb38-IN.

```
XX Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX Mycobacterium tuberculosis.
XX WO9942118-A2.
XX 26-AUG-1999.
XX 17-FEB-1999; 99WO-US03265.
XX 05-MAY-1998; 98US-0072596.
XX 18-FEB-1998; 98US-0024753.
XX (CORI-) CORIXA CORP.
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Iwardzik DR, Vedwick TS;
XX WPI; 1999-527416/44.
XX New polypeptide comprising antigenic portions of M. tuberculosis
PT Example 3; Page 179; 323pp; English.
XX This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.
XX Sequence 100 AA;
SQ
alignment_scores:
  Quality: 486.00 Length: 100
  Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000
alignment_block:
US-09-462-480-4 x AAY38993 ..
Align seg 1/1 to: AAY38993 from: 1 to: 100
1 ATGCAGAGATGAAGACCGATGCCGCTACCTCGGCGAGGAGCGAGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGACGGCAGGTTCGTTGCGAGCGGCGAGTGGCGCGCGCGGCGGACGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGGCGCGGTGGTGGCTTCCAAAGACGACCCCAATAGCAGCA 200
51 AlaGlnAlaAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAACCTCGACGAGATCTCGAGCAATATTCGTCAGCGCGGCTCCCAAT 250
67 nGluLeuAspLysIleSerThrAsnIleArgGlnAlaGlyValGlnTrp 84
251 CGAGGGCGGACGAGGACGAGCAGCGCGGTGCTCCTCGCAATGGGCTTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
seq_name: /SDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.AAB35218
```

```
seq_documentation_block:
ID AAB35218 standard; Protein; 100 AA.
XX AAB35218;
XX 24-APR-2001 (first entry)
XX M tuberculosis RV3874 protein.
XX Tuberculosis; TB; vaccine; esat-6 gene family; RV0287; RV1036C;
KW RV1037C; RV2346C; RV2348C; RV2653C; RV3020C; RV3444C;
KW RV3445C; RV3890C; RV3891C; RV3904C; RV3905C.
XX Mycobacterium tuberculosis.
XX WO200104151-A2.
XX 18-JAN-2001.
XX 13-JUL-2000; 2000WO-DK00398.
XX 13-JUL-1999; 99DK-0001020.
XX 15-JUL-1999; 99US-0144011.
XX (STAT-) STATENS SERUM INST.
XX Andersen P, Skjot R;
PI WPI; 2001-091923/10.
XX New polypeptide encoded by a member of the esat-6-gene family for
PT immunizing against and diagnosis of tuberculosis -
XX Example 2; Page 65; 80pp; English.
XX The present invention provides the protein and coding sequences for
CC members of the esat-6 gene family from Mycobacterium tuberculosis. These
CC proteins include RV0287, RV1036C, RV1037C, RV2346C, RV2348C, RV2653C,
CC RV2654C, RV3020C, RV3444C, RV3445C, RV3890C, RV3891C, RV3904C and
CC RV3905C. These can be used to produce vaccines against, and in the
CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of
CC the proteins of the invention.
XX Sequence 100 AA;
SQ
alignment_scores:
  Quality: 486.00 Length: 100
  Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000
alignment_block:
US-09-462-480-4 x AAB35218 ..
Align seg 1/1 to: AAB35218 from: 1 to: 100
1 ATGCAGAGATGAAGACCGATGCCGCTACCTCGGCGAGGAGCGAGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGACGGCAGGTTCGTTGCGAGCGGCGAGTGGCGCGCGCGGCGGACGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGGCGCGGTGGTGGCTTCCAAAGACGACCCCAATAGCAGCA 200
51 AlaGlnAlaAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAACCTCGACGAGATCTCGAGCAATATTCGTCAGCGCGGCTCCCAAT 250
```

|||||
67 ngluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGGCGGAGGAGGAGGAGGAGGCGTCTCGCAATGGGCTTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT: AAB19845

seq_documentation_block:
ID_ AAB19845 standard; Protein; 100 AA.
XX
AC AAB19845;
XX
DT 05-MAR-2001 (first entry)
DE Mycobacterium tuberculosis protein MTBN4.
XX
KW MTBN4; tuberculosis; BCG; vaccine; infection; diagnosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WC2000066157-A1.
XX
PD 09-NOV-2000.
XX
PF 04-MAY-2000; 2000WO-US12257.
XX
PR 04-MAY-1999; 99US-0132505.
XX
PA (PUBL-) PUBLIC HEALTH RES INST NEW YORK.
XX
PI Gennaro ML;
XX
DR WPI; 2001-007153/01.
DR N-PSDB; AAA89038.
XX
PT Novel polypeptide encoded by open reading frames present in
PT Mycobacterium tuberculosis genome and not by the BCG strain of M.
PT bovis, useful as vaccine and for diagnosing tuberculosis infection
XX
PS Claim 11; Fig 1; 35pp; English.
XX

The present sequence is that of the Mycobacterium tuberculosis
MTBN4 protein. This is 1 of 8 proteins, i.e. MTBN1-8 (see
AAB19842-49), encoded by 8 open reading frames (see AA89035-42)
identified as being present in the genome of M. tuberculosis but
absent from the genome of the BCG strain of Mycobacterium bovis.
MTBN1-8 represent reagents that are useful in discriminating between
M. tuberculosis and BCG and, in particular, for diagnostic methods
which discriminate between exposure of a subject to M. tuberculosis
and vaccination with BCG. The invention features these MTBN
polypeptides, functional fragments of them, DNA encoding them,
vectors, transformed cells, and diagnostic, therapeutic, and
prophylactic (vaccine) methods, including genetic vaccination
methods.
XX
SQ Sequence 100 AA;

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000
alignment_block:
US-09-462-480-4 x AAB19845 ..
Align seg 1/1 to: AAB19845 from: 1 to: 100
1 ATGGCAGAGATGACGCGTACCTCGGCGGAGGAGGAGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaThrLeuAlaGlnGluAlaGlyAs 17

51 TTTCGAGCGGATCTCCGCGACCTGAAACCAGATCGACCGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGAGCGGAGGTTCCTTCAGGGCCAGTGGCGGCGGCGGGGACGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGCGCGGTGGTGGCTTCCAAAGACGACGCAATAGCAGAAGCA 200
|||||
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAATCTGACGAGATCTCGACGAATATTGTCAGGCGCGCTCCAATACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGGCGGAGGAGGAGGAGGCGTCTCGCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT: AAW81746
seq_documentation_block:
ID_ AAW81746 standard; Protein; 802 AA.
XX
AC AAW81746;
XX
DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis fusion protein Tbf-2.
XX
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
OS Synthetic.
OS Mycobacterium tuberculosis.
XX
PN WC9816646-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18293.
XX
PR 13-MAR-1997; 97US-0818112.
PR 11-OCT-1996; 96US-0730510.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-261042/23.
DR N-PSDB; AAV64567.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
XX
PS Disclosure; Page 208-211; 230pp; English.
XX

This sequence represents the fusion protein Tbf-2 which is composed of
immunogenic polypeptides from Mycobacterium tuberculosis (MT). This
protein is used in a method for inducing protective immunity against
tuberculosis (TB). This sequence can be formulated into vaccines
and/or pharmaceutical compositions for immunising against
M. tuberculosis infection or may be used for the diagnosis of TB.
XX
SQ Sequence 802 AA;

alignment_scores:
Quality: 481.00 Length: 99

```

Ratio: 4.859          Gaps: 0
Percent Similarity: 100.000    Percent Identity: 98.990
alignment_block:
US-09-462-480-4 x AAW81746
..
Align seg 1/1 to: AAW81746 from: 1 to: 802

4 GCAGAGATGAAGACCGATGCCGCTACCTCGGGCAGGAGCGGTAATTT 53
425 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441
54 CGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGTGGAGTCGA 103
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerT 458
104 CGGCAGGTTCTGTCAGGCGCAGTGGCGCGCGCGGCGGAGCGGCCGCC 153
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474
154 CAGCGCGGCGTGGTGGCTTCCAGAACGAGCAGCAATAAGCAGACGAGCA 203
475 GlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 491
204 ACTCGACGAGATCTCGACGAATATTCGTACGCGCGCGCTCCAATACTCGA 253
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerA 508
254 GGGCGGACGAGCAGCAGCAGCGCGCTGCTCGCAATGGGCTTC 300
508 rgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 523

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64379
seq_documentation_block:
ID AAW64379 standard; Protein; 802 AA.
XX
AC AAW64379;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium antigen TbF2 protein fusion.
XX
KW Tuberculosis; infection; diagnosis; 38 kDa antigen; TbRa3; DPEP;
KW Tb38-1; TbF-2.
XX
OS Mycobacterium tuberculosis.
OS Synthetic.
XX
PN W09816645-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-251292/22.
DR N-PSDB; AAV55801.
XX
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
PS Example 7; Page 223-226; 250pp; English.
XX
CC This polypeptide comprises a fusion protein, designated TbF-2,
composed of Mycobacterium tuberculosis antigens TbRa3 (see AAW64295),
38 kDa antigen (see AAW64364), Tb38-1 (see AAW64321) and DPEP (see
AAW64322). It was produced by PCR amplification (see AAV4450-57) of
the appropriate antigen DNA sequences, cloning into an expression
vector, and expression in E. coli. TbF-2 can be used for
serodiagnosis of tuberculosis. The invention relates to
compositions and methods for diagnosing tuberculosis. It provides
polypeptides (see AAW64291-W64379) comprising antigenic or
immunogenic portions of M. tuberculosis antigens, or fusion proteins,
DNA sequences encoding such polypeptides, recombinant expression
vectors and host cells. Also claimed are methods and diagnostic
kits for detecting M. tuberculosis infection in a patient.
XX
SQ Sequence 802 AA;

alignment_scores:
Quality: 481.00          Length: 99
Ratio: 4.859            Gaps: 0
Percent Similarity: 100.000    Percent Identity: 98.990
alignment_block:
US-09-462-480-4 x AAW64379
..
Align seg 1/1 to: AAW64379 from: 1 to: 802

4 GCAGAGATGAAGACCGATGCCGCTACCTCGGGCAGGAGCGGTAATTT 53
425 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441
54 CAGCGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGTGGAGTCGA 103
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerT 458
104 CGGCAGGTTCTGTCAGGCGCAGTGGCGCGCGCGGCGGAGCGGCCGCC 153
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474
154 CAGCGCGGCGTGGTGGCTTCCAGAACGAGCAGCAATAAGCAGACGAGCA 203
475 GlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 491
204 ACTCGACGAGATCTCGACGAATATTCGTACGCGCGCGCTCCAATACTCGA 253
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerA 508
254 GGGCGGACGAGCAGCAGCAGCGCTGCTCGCAATGGGCTTC 300
508 rgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 523

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV32063
seq_documentation_block:
ID AAV32063 standard; Protein; 802 AA.
XX
AC AAV32063;
XX
DT 17-JAN-2000 (first entry)
XX
DE Mycobacterium tuberculosis antigen fusion protein TbF-2.
XX
KW Tuberculosis; antigen; fusion protein; TbF-2; TbRa3; 38kD; Tb38-1;
KW DPEP; diagnosis; therapy; vaccine; immunogen.
XX
OS Mycobacterium tuberculosis.
XX
PN W09951748-A2.
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999; 99WO-US07717.
XX
PR 07-APR-1998; 98US-0056556.
```

PR 30-DEC-1998; 98US-0223040.
XX (CORI-) CORIXA CORP.
XX Skeiky YAW, Alderson M, Campos-Neto A;
PI WPI; 1999-601610/51.
XX DR N-PSDB; AA220198.
XX New fusion proteins useful for diagnosis, prevention and treatment of
PT tuberculosis -
PT Claim 1; Fig 5G-J; 83pp; English.
XX
XX This sequence represents a recombinant Mycobacterium tuberculosis
CC tetra-antigen fusion protein, termed Tbf-2, composed of the antigens
CC Tba33, 39kp, Tba38-1 and DPEP. The fusion protein is expressed in
CC host cells using a vector carrying a polynucleotide (see AA220198)
CC comprising the 4 coding sequences. The invention provides fusion
CC proteins (see AAY32059-71) containing at least 2 M. tuberculosis
CC antigens. The new fusion proteins and polynucleotides encoding
CC them are useful as vaccines for preventing tuberculosis (claimed),
CC for diagnosis (via in vitro assays or intradermal skin tests for
CC detection of anti-M. tuberculosis antibodies), monitoring of
CC disease progression, and treatment of tuberculosis. They are more
CC effective immunogens than mixtures of the individual protein
CC components.
XX
SQ Sequence 802 AA;

alignment_scores:

Quality: 481.00 Length: 99
Ratio: 4.859 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.990

alignment_block:

US-09-462-480-4 x AAY32063 ..

Align seg 1/1 to: AAY32063 from: 1 to: 802

4 GCAGAGATGAGACCGATGCGGTACCTCGGCGAGGAGCGAGTAATTT 53
|||||
425 AlagluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441

54 CGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGAGTCGA 103
|||||
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSert 458

104 CGGAGGTTCGTTGTCAGGGCCAGTGGCGCGCGCGGCGGCGCGCGCC 153
|||||
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474

154 CAGGCCCGGTGGTGGCGCTTCCAAAGAGACGCCAATAAGCAGAGCAGGA 203
|||||
475 GlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysGlnGl 491

204 ACTCGAGAGATCTCGACGAATATTCGTACGGCGCGGTCCCAATACTCGA 253
|||||
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTrpSerA 508

254 GGGCCGACGAGGAGCAGCAGCGCGTGTCTCGCAAAATGGGCTTC 300
|||||
508 rgAlaAspGluGluGlnGlnAlaLeuSerSerGlnMetGlyPhe 523

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT: AAY39224

seq_documentation_block:

ID AAY39224 standard; Protein; 802 AA.

XX
AC AAY39224;

XX
DT 05-NOV-1999 (first entry)

XX M. tuberculosis fusion protein Tbf-6 amino acid sequence.
DE
XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
OS Synthetic.
XX Mycobacterium tuberculosis.
XX
PN WO9942076-A2.
XX
XX 26-AUG-1999.
XX
XX 17-FEB-1999; 99WO-US03268.
XX
XX 05-MAY-1998; 98US-0072967.
XX 18-FEB-1998; 98US-0025197.
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1999-527409/44.
DR N-PSDB; AA219457.
XX
XX New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
PT
XX Claim 37; Page 271-273; 299pp; English.
XX
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219457 to AA219460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.
XX
SQ Sequence 802 AA;

alignment_scores:

Quality: 481.00 Length: 99
Ratio: 4.859 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.990

alignment_block:

US-09-462-480-4 x AAY39224 ..

Align seg 1/1 to: AAY39224 from: 1 to: 802

4 GCAGAGATGAGACCGATGCGGTACCTCGGCGAGGAGCGAGTAATTT 53
|||||
425 AlagluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441

54 CGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGAGTCGA 103
|||||
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSert 458

104 CGGAGGTTCGTTGTCAGGGCCAGTGGCGCGCGCGGCGGCGCGCCGCC 153
|||||
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474

154 CAGGCCCGGTGGTGGCGCTTCCAAAGAGACGCCAATAAGCAGAGCAGGA 203
|||||
475 GlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysGlnGl 491

Quality: 481.00 Length: 99
Ratio: 4.859 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.990
alignment_block:
US-09-462-480-4 x AAY39081 ..
Align seg 1/1 to: AAY39081 from: 1 to: 802
4 GCAGAGATGAAGACCGCGCTACCGTCGGCGAGGAGCGAGTAATTT 53
|||||
425 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441
54 CGACCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGAGTCGA 103
|||||
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSert 458
104 CGGCGAGTTCGTCAGGGCCAGTGGCGCGCGCGGGGACGCGGCC 153
|||||
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474
154 CAGCCCGGTGGTGGCTTCCAAAGAGCAGCCCAATAAGCAGAAGCAGGA 203
|||||
475 GlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 491
204 ACTCGACGAGATCTCGACGAATATTCGTCAGGCGCGCTCCCAATACTCGA 253
|||||
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTrpSerA 508
254 GGGCGGAGGAGGAGCAGCAGCGCGTCTCGCAAAATGGGCTTC 300
|||||
508 rgAlaAspGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe 523

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT: AAY39033

seq_documentation_block:
ID AAY39033 standard; Protein; 802 AA.
XX
XX AAY39033;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis fusion protein TbF-2.
XX
KW Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX
OS Mycobacterium tuberculosis.
XX
PN W09942118-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999; 99NO-US03265.
XX
PR 05-MAY-1998; 98US-0072596.
PR 18-FEB-1998; 98US-0024753.
XX
XX (CORI-) CORIXA CORP.
PA
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;
XX
DR WPI: 1999-527416/44.
DR N-PSDB; AA219156.
XX
XX New polypeptide comprising antigenic portions of M. tuberculosis
XX
XX Example 10; Page 251-253; 323pp; English.
XX
XX This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a

CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.
XX
SQ Sequence 802 AA;

alignment_scores:
Quality: 481.00 Length: 99
Ratio: 4.859 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.990
alignment_block:
US-09-462-480-4 x AAY39033 ..
Align seg 1/1 to: AAY39033 from: 1 to: 802

4 GCAGAGATGAAGACCGATGCGCTACCGTCGGCGAGGAGCGAGTAATTT 53
|||||
425 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441
54 CGACCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGAGTCGA 103
|||||
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSert 458
104 CGGCGAGTTCGTCAGGGCCAGTGGCGCGCGCGGGGACGCGGCC 153
|||||
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474
154 CAGCCCGGTGGTGGCTTCCAAAGAGCAGCCCAATAAGCAGAAGCAGGA 203
|||||
475 GlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 491
204 ACTCGACGAGATCTCGACGAATATTCGTCAGGCGCGCTCCCAATACTCGA 253
|||||
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTrpSerA 508
254 GGGCGGAGGAGGAGCAGCAGCGCGTCTCGCAAAATGGGCTTC 300
|||||
508 rgAlaAspGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe 523

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT: AAU01901

seq_documentation_block:
ID AAU01901 standard; Protein; 983 AA.
XX
XX AAU01901;
XX
DT 29-AUG-2001 (first entry)
XX
DE M. tuberculosis TbF15 fusion protein.
XX
KW TbF15; antigen; vaccine; tuberculosis; AIDS; His tag; Ra3; 38kD;
KW 38-1; FL TbH4; acquired immunodeficiency disease.
XX
XX Synthetic.
OS Mycobacterium tuberculosis.
XX
XX
FH Key Location/Qualifiers
FT Binding-site 3..8 /label= Histidine_tag
FT /note= "Nickel chelating region used for purifying
FT the fusion protein"
FT 9..74
FT Region /label= Ra3_region
FT 75..425
FT Region /label= 38kD_region
FT 426..524
FT Region /label= 38-1_region
FT 525..983
FT Region /label= FL_TbH4_region
XX

```
PN WO200124820-A1.
XX
PD 12-APR-2001.
XX
PF 10-OCT-2000; 2000WO-US28095.
XX
PR 07-OCT-1999; 99US-0158338.
XX
PR 07-OCT-1999; 99US-0158425.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX WPI; 2001-290576/30.
DR N-PSDB; AAS03795.
XX
PT Vaccinating against Mycobacteria infections in mammals using fusion
PT proteins comprising combinations of heterologous antigens -
XX
PS Claim 6; Fig 4; 168pp; English.
XX
CC The sequence represents Mycobacterium fusion protein antigen Tbf15
CC consisting of a His tag for purification, antigens Ra3, 38KD, 38-1
CC and Fl-TbH4 (full-length Tbf4). Compositions comprising at least 2
CC heterologous antigens, as a fusion protein, and vectors expressing the
CC fusion proteins are used as vaccines to prophylactically immunise mammals
CC (especially humans) against infection by Mycobacteria. The compositions
CC contain at least 2 heterologous antigens that increase the serological
CC sensitivity of individuals infected with tuberculosis, a disease
CC frequently affecting patients with acquired immunodeficiency disease,
CC AIDS.
XX
SQ Sequence 983 AA;

alignment_scores:
    Quality: 481.00      Length: 99
    Ratio: 4.859        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 98.990

alignment_block:
US-09-462-480-4 x AAU01901 ..

Align seg 1/1 to: AAU01901 from: 1 to: 983

      4 GCAGAGATGAAGACCGATGCGCTACCTCGGCGAGGAGCGAGTAATT 53
      |||||
      426 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 442
      |||||
      54 CGAGCGGATCTCCGGCGACCTGAACCCAGATCGACCGAGTGAGTCGA 103
      |||||
      442 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSert 459
      |||||
      104 CGGCGAGTTCGTCAGGCGCAGTGGCGCGCGCGCGGCGGCGCGCGCC 153
      |||||
      459 hrAlaGlySerLeuGlnGlyGlnIleArgGlyAlaAlaGlyThrAlaAla 475
      |||||
      154 CAGCGCGCGGTGTCGCGCTTCCAAAGACGCCAATAAGCAGAACGACGGA 203
      |||||
      476 GlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysGlnGl 492
      |||||
      204 ACTCGACGAGATCTCGACGAATATTCGTACGGCGCGGTCCCAATACTCGA 253
      |||||
      492 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerA 509
      |||||
      254 GGGCGGACGAGGACGACGAGCGCTGCTCGCAAAATGGGCTTC 300
      |||||
      509 rgAlaAspGluGluGlnGlnGlnAlaLeuSerSerGlnMetGlyphe 524
```


Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-112-115	+	486.00	929.11	2.9e-44
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-111-110	+	486.00	929.11	2.9e-44
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-111-115	+	486.00	929.11	2.9e-44
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-056-556-115	+	486.00	929.11	2.9e-44
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-056-556-214	+	481.00	898.62	1.8e-43
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-112-88	+	462.00	883.56	1.1e-41
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-111-89	+	462.00	883.56	1.1e-41
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-056-556-88	+	462.00	883.56	1.1e-41
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-112-117	+	392.00	750.93	3.0e-34
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-111-112	+	392.00	750.93	3.0e-34
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-056-556-117	+	392.00	750.93	3.0e-34
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-112-98	+	122.00	243.26	1.7e-05
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-056-556-98	+	122.00	243.26	1.7e-05
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-112-96	+	121.00	241.70	2.1e-05
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-056-556-96	+	121.00	241.70	2.1e-05
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-112-95	+	118.00	235.94	4.4e-05
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-112-97	+	118.00	235.94	4.4e-05
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-056-556-95	+	118.00	235.94	4.4e-05
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-056-556-97	+	118.00	235.94	4.4e-05
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-112-93	+	114.00	227.90	0.0001
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-056-556-93	+	114.00	227.90	0.0001
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-220-538-116	-	98.50	178.37	0.0095
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-220-538-26	-	92.50	165.95	0.0425
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-220-528-32	-	92.50	165.20	0.0434
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-220-538-40	-	91.00	165.03	0.0582
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-199-637A-769	-	90.50	160.24	0.0734
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-413-814-78	-	88.50	144.33	0.1696
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-08-818-829-4	-	88.50	138.40	0.2010
/cgn2.5/ptodata/2/iaa/68 COMB	pep:US-09-199-637A-319	-	88.00	165.22	0.1026
/cgn2.5/ptodata/2/iaa/5A COMB	pep:US-07-626-618A-17	-	86.00	147.80	0.2477
/cgn2.5/ptodata/2/iaa/5A COMB	pep:US-08-086-439C-3	-	86.00	147.80	0.2477
/cgn2.5/ptodata/2/iaa/5A COMB	pep:US-08-056-051-2	-	86.00	147.80	0.2477
/cgn2.5/ptodata/2/iaa/5A COMB	pep:US-07-928-611-18	-	86.00	147.80	0.2477
/cgn2.5/ptodata/2/iaa/5A COMB	pep:US-08-333-977-17	-	86.00	147.80	0.2477
/cgn2.5/ptodata/2/iaa/5A COMB	pep:US-08-449-153-1	-	86.00	147.80	0.2477
/cgn2.5/ptodata/2/iaa/5A COMB	pep:US-08-434-877-3	-	86.00	147.80	0.2477
/cgn2.5/ptodata/2/iaa/5A COMB	pep:US-08-449-651-1	-	86.00	147.80	0.2477
/cgn2.5/ptodata/2/iaa/5A COMB	pep:US-08-818-811A-18	-	86.00	147.80	0.2477
/cgn2.5/ptodata/2/iaa/6A COMB	pep:US-08-833-703-1	-	86.00	147.80	0.2477
/cgn2.5/ptodata/2/iaa/6A COMB	pep:US-08-075-742-4	-	86.00	147.80	0.2477
/cgn2.5/ptodata/2/iaa/6A COMB	pep:US-09-060-694-18	-	86.00	147.80	0.2477

|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCAGGCGCGGTGCTCCAAAGCAGCAGCAATAAGCAGAACA 200
51 AlaGlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysG1 67
201 GGAACCTGCAGCAGATCTCGACGAATATTTCGTCAGGCGCGCTCCAATACT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGCGCAGCAGCAGCAGCAGCGCTGCTCCCAATGGCTTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-818-111-110

seq_documentation_block:
; Sequence 110, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-818-111-110

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-08-818-111-110 ..
Align seg 1/1 to: US-08-818-111-110 from: 1 to: 100

1 ATGCGAGATGAAGACCGATGCCGCTACCTCGGCGAGGAGGAGTAA 50

|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCAGCGGATCTCCGGCGACCTCAAACCCAGATCGACCAAGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGACGCGCAGGTTCGTTGACAGGCCAGTGGCGCGCGCGCGGGACGGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCAGGCGCGGTGCTCCAAAGCAGCAGCAATAAGCAGAACA 200
51 AlaGlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysG1 67
201 GGAACCTGCAGCAGATCTCGACGAATATTTCGTCAGGCGCGCTCCAATACT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGCGCAGCAGCAGCAGCAGCGCTGCTCCCAATGGCTTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-056-556-115

seq_documentation_block:
; Sequence 115, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-056-556-115

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-056-556-115 ..

Align seg 1/1 to: US-09-056-556-115 from: 1 to: 100.

```
1 ATGGCAGAGATGAAGACCGATGCGCTACCCCTCGGCGCAGGAGCAGGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTTCGAGCGGATCTCGGCGGACCTGAAACCCAGATCGACCGAGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGACGCGAGGTTCTGTCAGGGCCAGTGGCGGCGCGGCGGCGGAGCGGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCACGCGCGGCTGCTGCTTCCAGAGAGCAGCCCAATAGCAGAGCA 200
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAACTCGACGAGATCTCGACGAATATTCGTACAGCGCGCGTCCAATACT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGCGCGCAGGAGCAGCAGCAGCGCTGTCTCGCAAAATGGGCTTC 300
84 erArgAlaAspGluGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe 100
```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pap:US-09-056-556-214

seq_documentation_block:
; Sequence 214, Application US/09056556
; Patent No. 6350456

GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998

CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

alignment_scores:
Quality: 481.00
Ratio: 4.859

Length: 99
Gaps: 0

Percent Similarity: 100.000 Percent Identity: 98.990

alignment_block:

US-09-462-480-4 x US-09-056-556-214

Align seg 1/1 to: US-09-056-556-214 from: 1 to: 802

```
4 GCAGAGATGAAGACCGATGCGCTACCCCTCGGCGCAGGAGCAGGTAAATTT 53
|||||
425 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441
|||||
54 CGAGCGGATCTCGGCGACCTGAAACCCAGATCGACCGAGTGGAGTTCGA 103
|||||
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSert 458
|||||
104 CGCAGGTTCTGTCAGGGCCAGTGGCGGCGCGGCGGCGGAGCGGCC 153
|||||
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474
|||||
154 CAGGCGCGGCTGCTGCTTCCAGAGAGCAGCCCAATAGCAGAGCAGGA 203
|||||
475 GlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 491
|||||
204 ACTCGACGAGATCTCGACGAATATTCGTACAGCGCGCGTCCAATACTCGA 253
|||||
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerA 508
|||||
254 GGCGCGCAGGAGCAGCAGCAGCGCTGTCTCGCAAAATGGGCTTC 300
|||||
508 rgAlaAspGluGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe 523
```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pap:US-08-818-112-88

seq_documentation_block:
; Sequence 88, Application US/08818112
; Patent No. 6290969

GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids

REFLECTION NUMBER: 03/02

```
;
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-056-556-88

alignment_scores:
  Quality: 462.00      Length: 95
  Ratio: 4.863        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.947

alignment_block:
US-09-462-480-4 x US-09-056-556-88 ..
Align seg 1/1 to: US-09-056-556-88 from: 1 to: 95

16 ACCGATCCCGCTACCTCGGCGGAGGAGGTAATTTCGAGCGGATCTC 65
|||||
1 ThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPheGluArgIleSe 17
|||||
66 CGCGGACCTGAAACCCAGATCGACCAAGTGGAGTCGACGCGAGTTTCGT 115
|||||
17 rGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaGlySerL 34
|||||
116 TGCAGGCGCCAGTGGCGGCGCGCGGGGACGCGCCCGCCGCGCGGTG 165
|||||
34 euGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAlaAval 50
|||||
166 GTGGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAT 215
|||||
51 ValArgPheGlnGluAlaAlaAsnLysGlnLysGlnGluLeuAspGluI 67
|||||
216 CTGACCAATATTCTGTCAGGCGCGGCGGCGGCGGCGGCGGCGGCGG 265
|||||
67 eSerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAspGlu 84
|||||
266 AGCAGCAGCAGCGCGCTCTCTCCCAATGGGCTTC 300
|||||
84 luGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 95
|||||

seq_name: /cqn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-818-112-117

seq_documentation_block:
; Sequence 117, Application us/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
```

```
;
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 117:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-818-112-117

alignment_scores:
  Quality: 392.00      Length: 80
  Ratio: 4.900        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-480-4 x US-08-818-112-117 ..
Align seg 1/1 to: US-08-818-112-117 from: 1 to: 80

61 ATCTCCGGCGACCTGAAACCCAGATCGACCAAGTGGAGTCGACGCGCAGG 110
|||||
1 IleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaG 17
|||||
111 TTCTGTCAGGCGCAGTGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGCG 160
|||||
17 ySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAla 34
|||||
161 CGGTGGTGGCTTCCAAAGAGCAGCCCAATAGCAGACAGCAGGAACTCGAC 210
|||||
34 laValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnLysGlnLeu 50
|||||
211 GAGATCTCGACGAATATTCTGTCAGGCGCGGCTCCAATACTCGAGGCGCA 260
|||||
51 GluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAs 67
|||||
261 CGAGGAGCAGCAGCAGCGGTCTCTCGCAATGGGCTTC 300
|||||
67 pGluGluGlnGlnAlaLeuSerSerGlnMetGlyPhe 80
|||||

seq_name: /cqn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-818-111-112

seq_documentation_block:
; Sequence 112, Application us/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; TUBERCULOSIS
```

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 112:
SEQUENCE CHARACTERISTICS:
LENGTH: 80 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-112

alignment_scores:
Quality: 392.00 Length: 80
Ratio: 4.900 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-480-4 x US-08-818-111-112 ..
Align seg 1/1 to: US-08-818-111-112 from: 1 to: 80

61 ATCTCCGGCGACCTGAAACCCAGATCGACGAGTGGAGTCGACGGCAGG 110
|||||
1 IleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaG1 17
111 TTCGTTCCAGGGCCAGTGGCGCGCGGGGACGCGCGCCAGGCCG 160
|||||
17 YSerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAla 34
161 CGGTGGTGCCTTCCAAAGACGACCCCAATAAGCAGAGCAAGCACTGCAC 210
|||||
34 laValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnLeuAsp 50
211 GAGATCTCGACGATATTCGTCAGCGCGGTCCTCAATACCTCGAGGGCGGA 260
|||||
51 GluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAs 67
261 CGAGGACGACGACGCGCTGTCTCGCAAAATGGGCTTC 300
|||||
67 pGluGluGlnGlnAlaLeuSerSerGlnMetGlyPhe 80

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-056-556-117

seq_documentation_block:
; Sequence 117, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillion, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:

TREATM

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 117:
SEQUENCE CHARACTERISTICS:
LENGTH: 80 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-117

alignment_scores:
Quality: 392.00 Length: 80
Ratio: 4.900 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-480-4 x US-09-056-556-117 ..
Align seg 1/1 to: US-09-056-556-117 from: 1 to: 80

61 ATCTCCGGCGACCTGAAACCCAGATCGACGAGTGGAGTCGACGGCAGG 110
|||||
1 IleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaG1 17
111 TTCGTTCCAGGGCCAGTGGCGCGCGGGGACGCGCGCCAGGCCG 160
|||||
17 YSerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAla 34
161 CGGTGGTGCCTTCCAAAGACGACCCCAATAAGCAGAGCAAGCACTGCAC 210
|||||
34 laValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnLeuAsp 50
211 GAGATCTCGACGATATTCGTCAGCGCGGTCCTCAATACCTCGAGGGCGGA 260
|||||
51 GluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAs 67
261 CGAGGACGACGACGCGCTGTCTCGCAAAATGGGCTTC 300
|||||
67 pGluGluGlnGlnAlaLeuSerSerGlnMetGlyPhe 80

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-818-112-98

seq_documentation_block:
; Sequence 98, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillion, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.

```
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-98

alignment_scores:
  Quality: 122.00      Length: 25
  Ratio: 4.880        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-462-480-4 x US-08-818-112-98  ..

Align seg 1/1 to: US-08-818-112-98 from: 1 to: 28

226 ATTCGTCAGCGCGCTCCCAATCTCGAGCGCGGAGGAGGAGGAGGAGCA 275
|||||
4 IleArgGlnAlaGlyValGlnTyrSerArgAlaAspGluGlnGlnG1 20
|||||

276 GCGCTGCTCTCGCAATGGCTTC 300
|||||
20 nAlaLeuSerSerGlnMetGlyPhe 28

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-056-556-98

seq_documentation_block:
; Sequence 98, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-98

alignment_scores:
  Quality: 122.00      Length: 25
  Ratio: 4.880        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-462-480-4 x US-09-056-556-98  ..

Align seg 1/1 to: US-09-056-556-98 from: 1 to: 28

226 ATTCGTCAGCGCGCTCCCAATCTCGAGCGCGGAGGAGGAGGAGGAGCA 275
|||||
4 IleArgGlnAlaGlyValGlnTyrSerArgAlaAspGluGlnGlnG1 20
|||||

276 GCGCTGCTCTCGCAATGGCTTC 300
|||||
20 nAlaLeuSerSerGlnMetGlyPhe 28

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-818-112-96

seq_documentation_block:
; Sequence 96, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; COMPUTER READABLE FORM:
```

```
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-96

alignment_scores:
  Quality: 121.00      Length: 27
  Ratio: 4.481        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 92.593

alignment_block:
US-09-462-480-4 x US-08-818-112-96 ..
Align seg 1/1 to: US-08-818-112-96 from: 1 to: 27

133 GCGCGGGGGGCGCGCGCGCGCGGTGGTCCCAAGAAGC 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 GlycysglyThrAlaAlaGlnAlaAlaValArgPheGlnGluAl 17

183 AGCAATAAGCAGAGCAGCAACTCGACGAG 213
|||||:|||||:|||||:|||||:|||||:|||||:
17 aAlaAsnLysGlnLysGlnGluLeuAspGlu 27

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-056-556-96

seq_documentation_block:
; Sequence 96, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; TREATMENT OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-96

alignment_scores:
  Quality: 121.00      Length: 27
  Ratio: 4.481        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 92.593

alignment_block:
US-09-462-480-4 x US-09-056-556-96 ..
Align seg 1/1 to: US-09-056-556-96 from: 1 to: 27

133 GCGCGGGGGGCGCGCGCGCGCGGTGGTCCCAAGAAGC 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 GlycysglyThrAlaAlaGlnAlaAlaValArgPheGlnGluAl 17

183 AGCAATAAGCAGAGCAGCAACTCGACGAG 213
|||||:|||||:|||||:|||||:|||||:|||||:
17 aAlaAsnLysGlnLysGlnGluLeuAspGlu 27

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-818-112-95

seq_documentation_block:
; Sequence 95, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TREATMENT OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 95:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-95

alignment_scores:
  Quality: 118.00      Length: 24
```


Ratio: 4.917 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-462-480-4 x US-08-818-112-95 ..
Align seg 1/1 to: US-08-818-112-95 from: 1 to: 27
100 TCGACGGCAGGTTCTGTCAGGGCCAGTGGCGGCGCGGCGGGGACGGC 149
|||||
4 SerThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAl 20
150 CGCCCGAGCGCGGTGGTGGCGC 171
|||||
20 aAlaGlnAlaAlaValValArg 27

THIS PAGE BLANK (USPTO)

OM of: US-09-462-480-4 to: Pending_Patents_AA_Main:* out_format : pfs
Date: Jul 22, 2002 1:35 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+nt2p_model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09462480/runat_18072002_164419_19468/app_query.fasta_1.2850
-DB=Pending_Patents_AA_Main -QEMT=Fastan -SUFFIX=rapm
-CAPOP=12.000 -CAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosom62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODEL=LOCAL
-OUTPMT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480 -CGN1_1_271 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-462-480-4
Query length: 302
Database: Pending_Patents_AA_Main:*
Database sequences: 3502263
Database length: 351980561
Search time (sec): 542.380000

score_list:
Sequence Strd Orig zScore EScore Len i Documentation
/cgn2_6/ptodata/2/paa/US091_COMB.ppt:US-09-116-492-5 + 492.00 727.11 7.5e-32 100
/cgn2_6/ptodata/2/paa/US094_COMB.ppt:US-09-462-480-5 + 492.00 727.11 7.5e-32 100
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:PCT-US99-03265-110 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:PCT-US99-03265-115 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-729-622-110 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-730-510-115 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-341-110 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-578-115 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-024-753-110 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-025-197-115 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-596-110 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-596-115 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/US096_COMB.ppt:US-09-615-947-1 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-724-685-115 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/US100_COMB.ppt:US-10-084-843-115 + 486.00 718.32 2.3e-31 100
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:PCT-US99-03265-209 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:PCT-US99-03265-346 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:PCT-US99-03268-214 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:PCT-US99-03268-351 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-341-209 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-578-214 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-024-753-209 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-025-197-214 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-596-209 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-596-346 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-967-214 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-967-351 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-287-849-10 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/US100_COMB.ppt:US-10-084-843-214 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/US100_COMB.ppt:US-10-084-843-351 + 481.00 693.34 7.1e-31 802
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:PCT-US99-03265-89 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:PCT-US99-03268-88 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-658-800-89 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-659-683-88 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-680-573-89 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-680-574-88 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-729-622-88 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-730-510-88 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-341-89 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-578-88 + 462.00 683.63 2.1e-29 95

/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-024-753-89 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-025-197-88 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-596-89 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-967-88 + 462.00 683.63 2.1e-29 95
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-724-685-88 + 462.00 683.63 2.1e-29 95
seq_name: /cgn2_6/ptodata/2/paa/US091_COMB.ppt:US-09-116-492-5
seq_documentation_block:
; Sequence 5, Application US/09116492
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN
; TITLE OF INVENTION: FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY
; TITLE OF INVENTION: ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS METHODS USING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492
; EARLIER APPLICATION NUMBER: 60/052,631
; EARLIER FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-116-492-5
alignment_scores:
Quality: 492.00 Length: 100
Ratio: 4.920 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-462-480-4 x US-09-116-492-5
Align seg 1/1 to: US-09-116-492-5 from: 1 to: 100
1 ATGCACAGATGAACGACGATCGCGCTACCTCGCGCAGGAGGAGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlyGlnGluAlaGlyAs 17
|||||
51 TTTCGAGCGGATCTCCGCGCACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
|||||
101 CGACGGCAGGTTCGTTGCGAGCGGCAGTGGCGCGCGCGCGGACGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
|||||
151 GCCCAGCGCGGTTGGTGGCGCTTCCAAAGACAGCAGCAATAAGCAGAAGCA 200
|||||
51 AlaGlnAlaAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysGln 67
|||||
201 GGAACCTCGACGAGATCTCGAGCAATATTCGTACGCGCGCGGTCCAATACT 250
|||||
67 nGluLeuAspLutIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
|||||
251 CGAGGGCGGACGAGGACGAGCGAGCGGCTGTCTCCCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
|||||
seq_name: /cgn2_6/ptodata/2/paa/US094_COMB.ppt:US-09-462-480-5
seq_documentation_block:
; Sequence 5, Application US/09462480
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE

APPLICANT: BERTHET, FRANCIS-XAVIER
APPLICANT: ANDERSEN, PETER
APPLICANT: RASMUSSEN, PETER BIRK
TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS METHODS FOR THE PREPARATION OF SUCH FRAGMENTS
TITLE OF INVENTION: USING THE SAME
FILE REFERENCE: 0660-0165-0XPCT
CURRENT APPLICATION NUMBER: US/09/462,480
PRIOR FILING DATE: 2000-03-06
PRIOR APPLICATION NUMBER: PCT/IB98/01091
PRIOR FILING DATE: 1998-07-16
PRIOR APPLICATION NUMBER: 60/052,631
PRIOR FILING DATE: 1997-07-16
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 100
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-462-480-5

alignment_scores:
Quality: 492.00 Length: 100
Ratio: 4.920 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-462-480-4 x US-09-462-480-5
Align seg 1/1 to: US-09-462-480-5 from: 1 to: 100
1 ATGGCAGAGATGAAGACCGGATCGCGCTACCTCGGCGAGGAGCGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlyGlnGluAlaGlyAs 17
|||||
51 TTTCGACGGATCTCCGGGACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
|||||
101 CGACGGCAGGTTCTTCAGGCGCAGTGGCGCGCGGCGGAGCGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
|||||
151 GCCAGCCCGCGGTGTCGCTTCCAGAACGACGCCAATAAGCAGAGCA 200
|||||
51 AlaGlnAlaAlaValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
|||||
201 GGAACCTGACGAGATCTCGACGAAATATTCGTACGGCGCGGTCCAACTACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
|||||
251 CGAGGGCCGAGGAGGAGCAGCAGCGCGTCTCTCGCAAAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
|||||

seq_name: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:PCT-US99-03265-110

seq_documentation_block:

Sequence 110, Application PC/TUS9903265
GENERAL INFORMATION:
APPLICANT: Corixa Corporation
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds, LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US99/03265
FILING DATE: 17-FEB-1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/024,753
FILING DATE: 18-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 9532-0023-228
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 110:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US99-03265-110

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:

US-09-462-480-4 x PCT-US99-03265-110

Align seg 1/1 to: PCT-US99-03265-110 from: 1 to: 100

1 ATGGCAGAGATGAAGACCGGATCGCGCTACCTCGGCGAGGAGCGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
|||||
51 TTTCGACGGATCTCCGGGACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
|||||
101 CGACGGCAGGTTCTTCAGGCGCAGTGGCGCGCGGCGGAGCGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
|||||
151 GCCAGCCCGCGGTGTCGCTTCCAGAACGACGCCAATAAGCAGAGCA 200
|||||
51 AlaGlnAlaAlaValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
|||||
201 GGAACCTGACGAGATCTCGACGAAATATTCGTACGGCGCGGTCCAACTACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
|||||
251 CGAGGGCCGAGGAGGAGCAGCAGCGGTCTCTCGCAAAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
|||||

seq_name: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:PCT-US99-03268-115

seq_documentation_block:

Sequence 115, Application PC/TUS9903268
GENERAL INFORMATION:
APPLICANT: Corixa Corporation
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds, LLP

```

; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US99/03268
; FILING DATE: 17-FEB-1999
; CLASSIFICATION: 406
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/025,197
; FILING DATE: 18-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9532-0013-228
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
; PCT-US99-03268-115

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x PCT-US99-03268-115  ..
Align seg 1/1 to: PCT-US99-03268-115  from: 1 to: 100

1 ATGCGAGAGATGAAGACCGGTCGCGGTACCTCGGCGGAGGAGGAGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGAGCGGAGGTTCTGTCAGGCGCAGTGGCGCGCGCGCGGCGGAGCGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGGCGCGGTGGTGGCGCTTCCAGAAGCAGCCCAATAGCAGAAGCA 200
|||||
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAACCTCGACGAGATCTCGACGAATATTTCGTAGCGCGCGCTCCAATACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
|||||
84 erArgAlaAspGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 100
|||||
seq_name: /cgn_6/ptodata/2/paa/US087_COMB_pep:US-08-729-622-110
seq_documentation_block:
; Sequence 110, Application US/08729622
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas H.
; APPLICANT: Twardzik, David R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 132
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/729,622
; FILING DATE: 11-OCT-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
; US-08-729-622-110

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-08-729-622-110  ..
Align seg 1/1 to: US-08-729-622-110  from: 1 to: 100

1 ATGCGAGAGATGAAGACCGGTCGCGGTACCTCGGCGGAGGAGGAGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGAGCGGAGGTTCTGTCAGGCGCAGTGGCGCGCGCGCGGCGGAGCGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGGCGCGGTGGTGGCGCTTCCAGAAGCAGCCCAATAGCAGAAGCA 200
|||||
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAACCTCGACGAGATCTCGACGAATATTTCGTAGCGCGCGCTCCAATACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
|||||
84 erArgAlaAspGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 100
|||||
seq_name: /cgn_6/ptodata/2/paa/US087_COMB_pep:US-08-729-622-110
seq_documentation_block:
; Sequence 110, Application US/08729622
; GENERAL INFORMATION:
```

84 erArgAlaAspGluGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
|||||
seq_name: /cgn2_5/ptodata/2/paa/US087_COMB.pep:US-08-730-510-115
seq_documentation_block:
; Sequence 115, Application US/08730510
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 137
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/730,510
; FILING DATE: 27-AUG-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-730-510-115

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-08-730-510-115 ..

Align seg 1/1 to: US-08-730-510-115 from: 1 to: 100

```
1 ATGCGACGATGAGACCGGATCGCGCTACCTCGGCGAGGAGCGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
|||||
51 TTTCGACGGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
|||||
101 CGACGGCAGGTTCGTTCCAGGGCCAGTGGCGCGCGCGCGGCGGACGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
|||||
151 GCCAGGCGCGGTGGTGGCTTCCAGAGACGACCCCAATAGCAGACGA 200
|||||
```

51 AlaGlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
|||||
201 GGAACCTCAGCAGATCTCGACGAATATTCGTCAGGCGCGGTCCTCAACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
|||||
251 CGAGGGCGCAGCAGGAGCAGCAGCAGCGCTCTCTCGCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
|||||
seq_name: /cgn2_5/ptodata/2/paa/US089_COMB.pep:US-08-942-341-110
seq_documentation_block:
; Sequence 110, Application US/08942341
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,341
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-942-341-110

alignment_scores:

Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-08-942-341-110 ..

Align seg 1/1 to: US-08-942-341-110 from: 1 to: 100

```
1 ATGCGACGATGAGACCGGATCGCGCTACCTCGGCGAGGAGCGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
|||||
51 TTTCGACGGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
```

17 nPhleGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGAGCGAGGTTCTTCAGCGCCAGTGGCGCGCGCGCGGGGACGCGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCAGCGCGGTTGGTGGCGCTTCCAAAGAGCAGCAGCAATAAGCAAGCA 200
51 AlaGlnAlaAlaValAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAAGTCTGACGAGATCTCGAGCAATATTCGTACGCGCGCGGCTCCAACT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGCGCGGAGGAGCAGCAGCGCGTGTCTCGCAATGGGCTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe 100
seq_name: /cgn2_6/ptodata/2/paa/US089_COMB pep:US-08-942-578-115

seq_documentation_block:
; Sequence 115, Application US/08942578
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,578
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-942-578-115

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000
alignment_block:

US-09-462-480-4 x US-08-942-578-115
Align seg 1/1 to: US-08-942-578-115 from: 1 to: 100
1 ATGCCAGAGATGAAGACCGATGCGCTACCTCGGGCAGGAGGAGGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCTGAGCGGATCTCCGCGCACCTGAAACCAGATCGACATGACAGGTGAGT 100
17 nPhleGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGAGCGCAGGTTCTTCAGCGCCAGTGGCGCGCGCGGGGACGCGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCAGCGCGGTTGGTGGCGCTTCCAAAGAGCAGCAGCAATAAGCAAGCA 200
51 AlaGlnAlaAlaValAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAAGTCTGACGAGATCTCGAGCAATATTCGTACGCGCGCGGCTCCAACT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGCGCGGAGGAGCAGCAGCGGTGTCTCGCAATGGGCTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe 100
seq_name: /cgn2_6/ptodata/2/paa/US090_COMB pep:US-09-024-753-110

seq_documentation_block:
; Sequence 110, Application US/09024753
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 236
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/024,753
; FILING DATE: 18-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-024-753-110

```
alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-024-753-110 ..
Align seg 1/1 to: US-09-024-753-110 from: 1 to: 100

1 ATGCGAGAGATGAAGACCGATGCCCTACCTCGCGGAGGAGGAGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17

51 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCCAGCAGGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34

101 CGACGCGAGGTTCGTTCAGGCGCCAGTGGCGCGCGGCGGAGCGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50

151 GCCCAGCGCGGTGGTGGCTTCCAGAACGAGCCAAATAGCAGAACCA 200
|||||
51 AlaGlnAlaAlaValArgPheGlnGluAlaAlaAsnLysGlnLysG 67

201 GGAACCTCGACGAGATCTCGACGAATATTCGTACGCGCGGCTCCAACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTy 84

251 CGAGGCGCGAGGAGCAGCAGGCGGCTGTCTCGCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pap:US-09-025-197-115

seq_documentation_block:
; Sequence 115, Application US/09025197
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,197
; FILING DATE: 18-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C8
; TELECOMMUNICATION INFORMATION:
```

```
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-025-197-115

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-025-197-115 ..
Align seg 1/1 to: US-09-025-197-115 from: 1 to: 100

1 ATGCGAGAGATGAAGACCGATGCCCTACCTCGCGGAGGAGGAGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17

51 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCCAGCAGGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34

101 CGACGCGAGGTTCGTTCAGGCGCCAGTGGCGCGCGGCGGAGCGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50

151 GCCCAGCGCGGTGGTGGCTTCCAGAACGAGCCAAATAGCAGAACCA 200
|||||
51 AlaGlnAlaAlaValArgPheGlnGluAlaAlaAsnLysGlnLysG 67

201 GGAACCTCGACGAGATCTCGACGAATATTCGTACGCGCGGCTCCAACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTy 84

251 CGAGGCGCGAGGAGCAGCAGGCGGCTGTCTCGCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pap:US-09-072-596-110

seq_documentation_block:
; Sequence 110, Application US/09072596
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```



```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-09-072-596-110

alignment_scores:
    Quality: 486.00      Length: 100
    Ratio: 4.860        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-072-596-110 ..
Align seg 1/1 to: US-09-072-596-110 from: 1 to: 100

1 ATGGCAGAGATGAAGACCGATCGCGCTACCTCGGGCAGGAGGAGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACAGGTGGAGT 100
|||||
17 nPheGluArgLysSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGACGCGAGGTTCGTTGTCAGGCGCAGTGGCGCGCGCGGGGACGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGCGCGGGTGGTGGCTTCCAAAGCAGGCCAATAAGCAGAAGCA 200
|||||
51 AlaGlnAlaAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAATCGACGAGATCTCGACGAATATTCGTGAGCGCGCGCTCAATACT 250
|||||
67 nGluLeuAspGluLysSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
251 CGAGGCGCGAGCAGGAGCAGGAGCGGCTGCTCGCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/us090_COMB.pep:US-09-072-967-115

seq_documentation_block:
; Sequence 115, Application US/09072967
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP

;
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-09-072-967-115

alignment_scores:
    Quality: 486.00      Length: 100
    Ratio: 4.860        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-072-967-115 ..
Align seg 1/1 to: US-09-072-967-115 from: 1 to: 100

1 ATGGCAGAGATGAAGACCGATCGCGCTACCTCGGGCAGGAGGAGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACAGGTGGAGT 100
|||||
17 nPheGluArgLysSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGACGCGAGGTTCGTTGTCAGGCGCAGTGGCGCGCGCGGGGACGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGCGCGGGTGGTGGCTTCCAAAGCAGGCCAATAAGCAGAAGCA 200
|||||
51 AlaGlnAlaAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAATCGACGAGATCTCGACGAATATTCGTGAGCGCGCGCTCAATACT 250
|||||
67 nGluLeuAspGluLysSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
251 CGAGGCGCGAGCAGGAGCAGGAGCGGCTGCTCGCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/us090_COMB.pep:US-09-072-967-115

seq_documentation_block:
; Sequence 115, Application US/09072967
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP

;
; TITLE OF INVENTION: Tuberculosis vaccine and diagnostics
; TITLE OF INVENTION: based on the Mycobacterium tuberculosis esat-6 gene family
; FILE REFERENCE: 0459-0462P
```

```
; CURRENT APPLICATION NUMBER: US/09/615,947
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: fastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 100
; TYPE: PRT
; ORGANISM: M.Tuberculosis
US-09-615-947-1

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860         Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-615-947-1

Align seg 1/1 to: US-09-615-947-1 from: 1 to: 100

1 ATGCAGAGATGAAGACCGATCCGCTACCTCGGCGAGGCGAGGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACTGAAACCCAGATCGACCGAGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGAGCGCAGTTCGTTCGAGGCCAGTGGCGGGCGCGGGGAGCGCC 150
34 erThAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCAGCGCGCGTGTGCGGCTTCCAAGACGAGCCCAATAGCAAGCA 200
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAACTCGACGAGATCTCGACGAATATTGTCAGCGCGCGTCCCAATACT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGCGCGCAGCAGGAGCAGCAGCGCGCTGCTCCGCAATGGGCTTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-724-685-115

seq_documentation_block:
; Sequence 115, Application US/09724685
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir
; Dillon, Davin C.
; Campos-Neto, Antonio
; TITLE OF INVENTION: Compounds and Methods for
; Immunotherapy and Diagnosis of Tuberculosis
; NUMBER OF SEQUENCES: 155
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/724,685
; FILING DATE: 28-Nov-2000
; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/523,436
; FILING DATE: 01-SEP-1995
; APPLICATION NUMBER: US 08/533,634
; FILING DATE: 22-SEP-1995
; APPLICATION NUMBER: US 08/620,874
; FILING DATE: 22-MAR-1996
; APPLICATION NUMBER: US 08/659,683
; FILING DATE: 05-JUN-1996
; APPLICATION NUMBER: US 08/680,574
; FILING DATE: 12-JUL-1996
; APPLICATION NUMBER: WO PCT/US96/14674
; FILING DATE: 30-AUG-1996
; APPLICATION NUMBER: US 08/730,511
; FILING DATE: 11-OCT-1996
; APPLICATION NUMBER: US 08/818,112
; FILING DATE: 13-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 014058-008561US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 115:
US-09-724-685-115

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860         Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-724-685-115

Align seg 1/1 to: US-09-724-685-115 from: 1 to: 100

1 ATGCAGAGATGAAGACCGATCCGCTACCTCGGCGAGGCGAGGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACTGAAACCCAGATCGACCGAGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGAGCGCAGTTCGTTCGAGGCCAGTGGCGGGCGCGGGGAGCGCC 150
34 erThAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCAGCGCGCGTGTGCGGCTTCCAAGACGAGCCCAATAGCAAGCA 200
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAACTCGACGAGATCTCGACGAATATTGTCAGCGCGCGTCCCAATACT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGCGCGCAGCAGGAGCAGCAGCGCGCTGCTCCGCAATGGGCTTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-724-685-115

seq_documentation_block:
; Sequence 115, Application US/09724685
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir
; Dillon, Davin C.
; Campos-Neto, Antonio
; TITLE OF INVENTION: Compounds and Methods for
; Immunotherapy and Diagnosis of Tuberculosis
; NUMBER OF SEQUENCES: 155
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/724,685
; FILING DATE: 28-Nov-2000
; CLASSIFICATION: <Unknown>
```

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 115:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 115:
US-10-084-843-115

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000
alignment_block:
US-09-462-480-4 x US-10-084-843-115
Align seq 1/1 to: US-10-084-843-115 from: 1 to: 100
1 ATGCAGAGTGTGCGGACCTGAGGAGTATTCGTCAGGCGGCGTCCAATACT 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGACCTGAAACCCAGATCGACCGAGGTGGAGT 100
|||||
17 nPheGluArgLysSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGACGCGAGTTCGTCAGGCGCAGTGGCGGCGGCGGCGGCGGCGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGCGCGGTGGTGGCTTCCAGAGACGACGCCAATAGCAGAGCA 200
|||||
51 AlaGlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67

201 GGAACCTCGACGAGATCTCGACGAATATTCGTCAGGCGGCGTCCAATACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGGCGGACGAGGAGGAGCAGCAGGCGGTCTCTCGCAAATGGGCTTC 300
|||||
84 erArgAlaAspGluGluGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

OM of: US-09-462-480-4 to: Pending_Patents_AA_New:* out_format : pfs
Date: Jul 22, 2002 1:37 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+n2p.spool/US09462480/runat_18072002_164419_19498/app_query.fasta_1.2850
-DB=Pending_Patents_AA_New -QFMT=fastan -SUFFIX=rapn
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480_SCGN1_1_93 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-462-480-4
Query length: 302
Database: Pending_Patents_AA_New:*
Database sequences: 378952
Database length: 124292526
Search time (sec): 104.950000

score_list:	Strd Orig	zScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-5 +	492.00	756.62	6.1e-34	10	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-5 +	492.00	756.62	6.1e-34	100	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-5 +	492.00	756.62	6.1e-34	100	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-639 +	486.00	747.47	2.0e-33	49	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-09-116-492A-6 +	247.00	389.04	3.7e-13	49	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-7 +	247.00	389.04	3.7e-13	49	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-09-116-492A-6 +	205.00	326.30	1.4e-09	42	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-7 +	205.00	326.30	1.4e-09	42	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-09-116-492A-28 +	187.00	291.40	5.0e-08	10	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-28 +	187.00	291.40	5.0e-08	10	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-12 +	187.00	291.40	5.0e-08	10	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-09-116-492A-12 +	137.00	226.05	0.0008	2	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-12 +	137.00	226.05	0.0008	28	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17077 -	119.00	169.55	0.0369		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17076 -	119.00	169.35	0.0370		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17480 -	119.00	169.23	0.0371		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17479 -	119.00	169.03	0.0372		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17075 -	119.00	168.91	0.0372		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17478 -	119.00	168.80	0.0373		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-71465 +	109.50	167.32	0.2052		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-26387 +	102.50	146.90	0.9045		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-26077 +	100.50	148.07	1.27		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-09-116-492A-8 +	100.00	172.50	1.05	20	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-8 +	100.00	172.50	1.05	20	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-10-155-881-10947 +	99.00	154.59	1.54		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-60-389-987-2231 +	98.50	129.65	2.25		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10418 -	96.50	147.24	2.62		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-9649 -	96.00	149.52	2.79		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-9768 +	96.00	144.33	2.96		
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2448 -	96.00	132.37	3.40		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-11792 +	95.50	140.65	3.38		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-11792 -	95.50	140.65	3.38		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-17185 -	95.00	141.45	3.66		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10185 -	95.00	137.68	3.82		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-7135 -	94.50	149.11	3.66		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10473 -	94.50	142.69	3.94		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-27136 -	93.50	141.94	4.75		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-17173 +	93.00	140.95	5.25		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-12231 -	93.00	139.43	5.34		
/cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:PCT-US02-06388-2 -	92.50	140.49	5.77	220	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2409 -	92.50	140.11	5.79		

/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-11043 - 91.50 138.44 7.06
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-26985 + 91.50 136.48 7.22
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-11656 + 91.00 135.80 7.96
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-32863 - 90.50 137.48 8.53
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10047 - 90.50 133.91 8.89

seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-5

seq_documentation_block:
; Sequence 5, Application US/09116492A
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM M
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, I
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492A
; CURRENT FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-116-492A-5

alignment_scores:
Quality: 492.00 Length: 100
Ratio: 4.920 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-480-4 x US-09-116-492A-5

Align seg 1/1 to: US-09-116-492A-5 from: 1 to: 100

1 ATGCAGAGATGAACGCGATCGCGCTACCTCGGCGCAGGAGGAGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaThrLeuGlyGlnGluAlaGlyAs 17

51 TTTTCGAGGATCTCCGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34

101 CGACGCGAGGTTCGTCAGGCGCAGTGGCGCGCGCGGCGGCGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnIleArgGlyAlaAlaGlyThrAla 50

151 GCCCAGCGCGCGGTGGTGGCTTCCAAAGAACGACGCCAATAAGCAAGCA 200
|||||
51 AlaGlnAlaAlaValAlaValArgPheGlnGluAlaAlaAsnLysGlnLysG 67

201 GGAACCTGACGAGATCTCGACGATATTCGTCAGCGCGCGCTCCAACTACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnIleYrs 84

251 CGAGCGCGCGCAGCAGCAGCAGCGGCTGCTCGCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-5

seq_documentation_block:
; Sequence 5, Application US/10140045
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER

```
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/10/140,045
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: US/09/116,492A
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-10-140-045-5

alignment_scores:
  Quality: 492.00      Length: 100
  Ratio: 4.920        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-462-480-4 x US-10-140-045-5 ..
Align seg 1/1 to: US-10-140-045-5 from: 1 to: 100

1 ATGCGAGAGATGAAGACCGATGCCCTACCTCGGCGAGGAGGAGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlyGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGCGACCTGAAACCCAGATCGACGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGAGCGCAGGTTCTGTCAGGCGCAGTGGCGCGCGCGGCGGCGGCGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCGAGCGCGGTTGTCGCGCTTCCAAAGCAGCCCAATAAGCAGCAACA 200
|||||
51 AlaGlnAlaAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAAGTCTGACGAGATCTCGACGAATATTCGTACGCGCGCGCTCCAACTACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
151 CGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
251 CGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/us10_NEW_COMB.pap:us-10-080-170B-639

seq_documentation_block:
; Sequence 639, Application US/10080170B
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170B
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 639
; LENGTH: 100
```

```
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-10-080-170B-639

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-10-080-170B-639 ..
Align seg 1/1 to: US-10-080-170B-639 from: 1 to: 100

1 ATGCGAGAGATGAAGACCGATGCCCTACCTCGGCGAGGAGGAGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGCGACCTGAAACCCAGATCGACGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGAGCGCAGGTTCTGTCAGGCGCAGTGGCGCGCGCGGCGGCGGCGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCGAGCGCGGTTGTCGCGCTTCCAAAGCAGCCCAATAAGCAGCAACA 200
|||||
51 AlaGlnAlaAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAAGTCTGACGAGATCTCGACGAATATTCGTACGCGCGCGCTCCAACTACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/us09_NEW_COMB.pap:us-09-116-492A-6

seq_documentation_block:
; Sequence 6, Application US/09116492A
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492A
; CURRENT FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-116-492A-6

alignment_scores:
  Quality: 247.00      Length: 49
  Ratio: 5.041        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-462-480-4 x US-09-116-492A-6 ..
```

;; APPLICANT: RASMUSSEN, PETER B

```
US-09-462-480-4 x US-10-140-045-7
Align seg 1/1 to: US-10-140-045-7 from: 1 to: 42

175 CAAGAAGCAGCAATAGCAGACAGCAAGTCCGACGAGATCTCGACGAA 224
|||||
1 GlnGluAlaAlaAsnLysGlnLysGlnGluLeuAspGluIleSerThrAs 17
|||||
225 TATTCGTCAGCGCGGCTCCCAATACTCGAGGGCGGACGAGGACGAGC 274
|||||
17 nIleArgGlnAlaGlyValGlnTrpSerArgAlaAspGluGlnGlnG 34
|||||
275 AGGCGCTGTCTCCCAATAGGCTTC 300
|||||
34 InAlaLeuSerSerGlnMetGlyPhe 42

seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-28

seq_documentation_block:
; Sequence 28, Application US/09116492A
; GENERAL INFORMATION:
; APPLICANT: GICOUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM LEPRAE, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR THERAPY OF TUBERCULOSIS
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492A
; CURRENT FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-09-116-492A-28
```

```
alignment_scores:
  Quality: 187.00      Length: 100
  Ratio: 2.367        Gaps: 0
  Percent Similarity: 79.000  Percent Identity: 40.000

alignment_block:
US-09-462-480-4 x US-09-116-492A-28

Align seg 1/1 to: US-09-116-492A-28 from: 1 to: 100

1 ATGCGACAGATGAAGACCGGATCGCGCTACCTCGGCGAGGAGCGAGTAA 50
|||||
1 MetAlaGluMetIleThrGluAlaAlaIleLeuThrGlnGlnAlaIaGl 17
|||||
51 TTTCGAGCGGATCTCCGCGCACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheAspGlnIleAlaSerGlyLeuSerGlnGluArgAsnPheValAspS 34
|||||
101 CGACGCGAGGTTCTGTCAGGCGCAGTGGCGCGCGCGGCGGAGCGGCC 150
|||
34 erIleGlyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAla 50
|||||
151 GCCCAGGCGCGGTTGGTTCGCTTCCAGAGACGACGACCAATAGCAGCA 200
|||
51 AlalaLeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleAr 67
|||||
201 GGAATCGACGAGATCTCGACGAATATTCGTCAGGCGCGCTCCCAATACT 250
|||||
67 glnLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnTyrT 84
|||||
251 CGAGGCGCGAGGAGCAGCAGGCGGTGCTCTCGCAATGGGCTTC 300
|||
34 erIleGlyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAla 50
|||||
151 GCCCAGGCGCGGTTGGTTCGCTTCCAGAGACGACGACCAATAGCAGCA 200
|||
51 AlalaLeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleAr 67
|||||
201 GGAATCGACGAGATCTCGACGAATATTCGTCAGGCGCGCTCCCAATACT 250
|||||
67 glnLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnTyrT 84
|||||
251 CGAGGCGCGAGGAGCAGCAGGCGGTGCTCTCGCAATGGGCTTC 300
```

```
seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-28

seq_documentation_block:
; Sequence 28, Application US/10140045
; GENERAL INFORMATION:
; APPLICANT: GICOUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM LEPRAE, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR THERAPY OF TUBERCULOSIS
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/10/140,045
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: US/09/116,492A
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-140-045-28

alignment_scores:
  Quality: 187.00      Length: 100
  Ratio: 2.367        Gaps: 0
  Percent Similarity: 79.000  Percent Identity: 40.000

alignment_block:
US-09-462-480-4 x US-10-140-045-28

Align seg 1/1 to: US-10-140-045-28 from: 1 to: 100

1 ATGCGACAGATGAAGACCGGATCGCGCTACCTCGGCGAGGAGCGAGTAA 50
|||||
1 MetAlaGluMetIleThrGluAlaAlaIleLeuThrGlnGlnAlaIaGl 17
|||||
51 TTTCGAGCGGATCTCCGCGCACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheAspGlnIleAlaSerGlyLeuSerGlnGluArgAsnPheValAspS 34
|||||
101 CGACGCGAGGTTCTGTCAGGCGCAGTGGCGCGCGCGGCGGAGCGGCC 150
|||
34 erIleGlyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAla 50
|||||
151 GCCCAGGCGCGGTTGGTTCGCTTCCAGAGACGACGACCAATAGCAGCA 200
|||
51 AlalaLeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleAr 67
|||||
201 GGAATCGACGAGATCTCGACGAATATTCGTCAGGCGCGCTCCCAATACT 250
|||||
67 glnLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnTyrT 84
|||||
251 CGAGGCGCGAGGAGCAGCAGGCGGTGCTCTCGCAATGGGCTTC 300
|||
84 hrLysThrAspAspGluAlaAsnGlnLeuLeuSerSerLysMetAsnPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-12

seq_documentation_block:
; Sequence 12, Application US/10080170B
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR THERAPY OF TUBERCULOSIS
```



```
; TITLE OF INVENTION: MODULATING VARIOUS RESPONSES
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 17077
; LENGTH: 827
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..827
; OTHER INFORMATION: Ceres Seq. ID no. 2708208
US-09-935-625-17077
```

```
alignment_scores:
  Quality: 119.00      Length: 111
  Ratio: 2.052         Gaps: 6
  Percent Similarity: 52.252      Percent Identity: 35.135
```

```
alignment_block:
```

```
US-09-462-480-4/rev x US-09-935-625-17077 ..
```

```
Align seg 1/1 to: US-09-935-625-17077 from: 1 to: 827
```

```
301 AGAAGCCATTGCGAGGACAGCGCTGCTGCTCTCT...CGTCGGCC 255
|||||.....|.....|.....|.....|.....|.....|
361 ArgSerProLeuAlaLysArgGlyArgSerAspSerProGlyArgSerPr 377
|||||.....|.....|.....|.....|.....|.....|
254 CTCGAGTATTGGACGCGGCTGACGAATATTCTCGAGA..... 215
|||||.....|.....|.....|.....|.....|.....|
377 oSerProValAlaArgLeuArgAspProThrGlyAlaArgLeuProSerP 394
|||||.....|.....|.....|.....|.....|.....|
214 .....TCTCGTCGAGTTCCTGCTTCTGCTTATTGGCTGCTCTTTGGA 173
|||||.....|.....|.....|.....|.....|.....|
394 roSerIleGluGlnArgLeuProSerProProValAlaGlnArgLeuPro 410
|||||.....|.....|.....|.....|.....|.....|
172 AGCGCACACCG...CGGCTGGCGGCGCTCCCGCGCGCGCGCCGCAC 126
|||||.....|.....|.....|.....|.....|.....|
411 SerProProProArgArgAlaGlyLeuProSerProPro..... 423
|||||.....|.....|.....|.....|.....|.....|
125 TGGCCCTGCAACGACCTGCGTGCACCTCCACCTGGTCGATCTGGGTTT 76
|||||.....|.....|.....|.....|.....|.....|
424 ...ProAlaGlnArgLeuProSerProProProArgArgAlaGlyLeup 439
|||||.....|.....|.....|.....|.....|.....|
75 CAGGTCCCGCGAGA.....TCCGCTCGAAATTACCTG..... 44
|||||.....|.....|.....|.....|.....|.....|
439 roSerProMetArgIleGlyGlySerHisAlaAlaAsnHisLeuGluSer 455
|||||.....|.....|.....|.....|.....|.....|
43 .....CCTCCTGCCCGCA 32
|||||.....|.....|.....|.....|.....|.....|
456 ProSerProSerSerLeuSerProProGlyArg 466
```

```
seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17076
```

```
seq_documentation_block:
; Sequence 17076, Application US/099335625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 17076
; LENGTH: 847
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..847
```

```
; OTHER INFORMATION: Ceres Seq. ID no. 2708207
US-09-935-625-17076
```

```
alignment_scores:
  Quality: 119.00      Length: 111
  Ratio: 2.052         Gaps: 6
  Percent Similarity: 52.252      Percent Identity: 35.135
```

```
alignment_block:
```

```
US-09-462-480-4/rev x US-09-935-625-17076 ..
```

```
Align seg 1/1 to: US-09-935-625-17076 from: 1 to: 847
```

```
301 AGAAGCCATTGCGAGGACAGCGCTGCTGCTCTCT...CGTCGGCC 255
|||||.....|.....|.....|.....|.....|.....|
381 ArgSerProLeuAlaLysArgGlyArgSerAspSerProGlyArgSerPr 397
|||||.....|.....|.....|.....|.....|.....|
254 CTCGAGTATTGGACGCGGCTGACGAATATTCTCGAGA..... 215
|||||.....|.....|.....|.....|.....|.....|
397 oSerProValAlaArgLeuArgAspProThrGlyAlaArgLeuProSerP 414
|||||.....|.....|.....|.....|.....|.....|
214 .....TCTCGTCGAGTTCCTGCTTCTGCTTATTGGCTGCTCTTTGGA 173
|||||.....|.....|.....|.....|.....|.....|
414 roSerIleGluGlnArgLeuProSerProProValAlaGlnArgLeuPro 430
|||||.....|.....|.....|.....|.....|.....|
172 AGCGCACACCG...CGGCTGGCGGCGCTCCCGCGCGCGCGCCGCAC 126
|||||.....|.....|.....|.....|.....|.....|
431 SerProProProArgArgAlaGlyLeuProSerProPro..... 443
|||||.....|.....|.....|.....|.....|.....|
125 TGGCCCTGCAACGACCTGCGTGCACCTCCACCTGGTCGATCTGGGTTT 76
|||||.....|.....|.....|.....|.....|.....|
444 ...ProAlaGlnArgLeuProSerProProProArgArgAlaGlyLeup 459
|||||.....|.....|.....|.....|.....|.....|
75 CAGGTCCCGCGAGA.....TCCGCTCGAAATTACCTG..... 44
|||||.....|.....|.....|.....|.....|.....|
459 roSerProMetArgIleGlyGlySerHisAlaAlaAsnHisLeuGluSer 475
|||||.....|.....|.....|.....|.....|.....|
43 .....CCTCCTGCCCGCA 32
|||||.....|.....|.....|.....|.....|.....|
476 ProSerProSerSerLeuSerProProGlyArg 486
```

```
seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17480
```

```
seq_documentation_block:
; Sequence 17480, Application US/099335625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPA
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 17480
; LENGTH: 859
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..859
; OTHER INFORMATION: Ceres Seq. ID no. 3023746
US-09-935-625-17480
```

```
alignment_scores:
  Quality: 119.00      Length: 111
  Ratio: 2.052         Gaps: 6
  Percent Similarity: 52.252      Percent Identity: 35.135
```

```
alignment_block:
```

```
US-09-462-480-4/rev x US-09-935-625-17480 ..
```

```

Align seg 1/1 to: US-09-935-625-17480 from: 1 to: 859
301 AGAAGCCCATTTGGAGAGCAGCGCTGCTGCTCT...CGTCGGCC 255
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
393 ArgSerProLeuAlaLysArgGlyArgSerProGlyArgSerPr 409
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
254 CTCGAGTATTGGACGCGCGCTGACGAATATTCGTCGAGA..... 215
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
409 oSerProValAlaArgLeuArgAspProThrGlyAlaArgLeuProSerP 426
214 .....TCTGTCGAGTTCCTCTCTCTCTATTGCTGCTCTTTGGA 173
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
426 roSerIleGluInArgLeuProSerProProValAlaGlnArgLeuPro 442
172 AGCGCACACCG...CGGCTGGGCGCGCTCCCGCGCGCGCGCGCAC 126
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
443 SerProProProArgArgAlaGlyLeuProSerProPro..... 455
125 TGGCCCTGTCAACGACCTGCGTCCGACTCCACCTGTCGATCTGGGTTTT 76
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
456 ....ProAlaGlnArgLeuProSerProProProArgArgAlaGlyLeuP 471
75 CAGGTCCCGCGAGA.....TCCGCTCGAAATTACCTG..... 44
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
471 roSerProMetArgIleGlySerHisAlaAlaAsnHisLeuGluSer 487
43 .....CCTCCTGCCCGA 32
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
488 ProSerProSerSerLeuSerProProGlyArg 498

```

THIS PAGE BLANK (USPTO)

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
p1r2:H70802	+	486.00	844.89	5.7e-39	100	hypothetical protein Rv3874 - M
p1r2:T10032	+	187.00	326.15	4.5e-10	100	hypothetical protein MCBG28.13
p1r2:G84693	+	105.50	190.60	0.0018	891	probable proline-rich protein
p1r2:AA0671	-	119.50	167.43	0.0359	864	dynam, internal form 1, long
p1r1:EDBEIF	-	104.00	160.61	0.0510	1460	immediate-early protein TE180
p1r2:T35785	+	99.50	157.06	0.1365	859	probable beta-glucosidase - Str
p1r2:B75514	+	98.50	156.93	0.1366	873	penicillin-binding protein 1 -
p1r2:T18535	+	99.00	153.84	0.1540	1151	high molecular mass nuclear an
p1r2:T16953	+	97.50	164.99	0.2039	208	hypothetical protein T28D9.2 -
p1r2:S11508	+	97.50	153.67	0.2130	851	D100 protein - rat
p1r2:BA0671	-	97.50	153.67	0.2130	851	dynam, internal form 2, short
p1r2:T34871	-	96.50	156.06	0.2619	509	probable membrane protein - Str
p1r2:H72706	+	95.00	165.28	0.3494	117	hypothetical protein APE1071 -
p1r1:QR8072	-	94.50	154.49	0.4057	402	microtubule-associated protein
p1r2:T18918	+	94.00	159.36	0.4436	197	hypothetical protein C04G2.8 -
p1r2:G86301	+	94.00	153.39	0.4539	194	arginine/serine-rich protein
p1r2:J00431	-	93.50	153.92	0.5046	348	hypothetical 35.5K protein - Str
p1r1:AA5344	-	93.50	142.47	0.5273	1446	immediate-early protein - sul
p1r2:S31719	-	93.00	152.36	0.5654	379	proline-rich protein - African
p1r2:S33121	-	93.00	144.77	0.5822	975	homeotic protein CDP - dog (fra
p1r2:T06174	-	92.50	160.16	0.6113	129	PZE40 protein - barley (fragment
p1r2:F75518	-	92.50	145.11	0.6477	839	hypothetical protein - Deinococ
p1r2:T35661	+	92.50	142.33	0.6547	1186	probable chromosome associated
p1r2:AH2743	-	92.00	149.94	0.7082	413	conserved hypothetical protein
p1r2:G97524	+	92.00	149.88	0.7084	416	hypothetical protein AGR_C2517
p1r2:B99103	+	91.50	149.33	0.7914	410	hypothetical protein pXOI-98 -
p1r2:AS5301	-	91.50	144.46	0.8057	733	microtubule-associated protein
p1r2:JS0170	-	90.50	150.37	0.9773	283	collagen col-19 - Caenorhabdit
p1r2:T27708	-	90.50	150.20	0.9780	289	hypothetical protein ZK1193.1 -
p1r2:T02345	+	90.50	135.54	1.0303	1791	hypothetical protein KIAA0324
p1r2:T39843	+	90.00	141.67	1.13	750	coiled coil protein - fission y
p1r2:B45219	+	89.00	137.57	1.14	1250	N-methyl-D-aspartate receptor
p1r2:T36632	+	89.00	145.74	1.23	406	probable oxidoreductase - Strept
p1r2:SA1495	-	89.50	144.68	1.24	463	dc stretch-binding protein CSB
p1r2:SA4363	-	89.50	144.66	1.24	464	transformation upregulated nucl
p1r2:AS4143	-	89.50	144.66	1.24	464	kappa-B motif-binding phosphor
p1r2:AS5197	+	89.50	144.03	1.24	502	Wiskott-Aldrich syndrome prote
p1r2:B70600	+	89.50	145.90	1.32	103	hypothetical protein RV3905c -
p1r2:S16506	+	88.50	144.51	1.54	381	hypothetical protein - human
p1r2:TA4252	+	88.00	144.08	1.72	361	splicing factor-like protein -

A;Title: Use of an ordered cosmid library to deduce the genomic organization of Mycobact
A;Reference number: Z16917; MUID:93188700
A;Accession: T10032
A;Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-100 <IG>
A;Cross-references: EMBL:Y14967; NID:g2370268; PIDN:CAA75210.1; PID:g2370280

alignment_scores:
Quality: 187.00 Length: 100
Ratio: 2.367 Gaps: 0
Percent Similarity: 79.000 Percent Identity: 40.000

alignment_block:

US-09-462-480-4 x T10032 ..

Align seg 1/1 to: T10032 from: 1 to: 100

```
1 ATGCAGAGATGAGACGATGCGCTACCTCGGCGAGGAGGTAA 50
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1 MetAlaGluMetIleThrGluAlaLeuThrGlnGlnAlaAlaG1 17

51 TTTCAGCGGATCTCCGGCGACCTGAACCCAGATCGACCAAGTGGAGT 100
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
17 nPheAspGlnIleAlaSerGlyLeuSerGlnGluArgAsnPheValAsps 34

101 CGAGCGGAGGTCTGTCGAGCGCAGTGGCGCGCGGGCGGAGCGCC 150
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
34 erIleGlyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAla 50

151 GCCAGCGCGCGGTGGTGGCTTCCAAAGACGACCCAAATAGCAGAAGCA 200
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
51 AlaLeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleAr 67

201 GGAATCGACGAGATCGAGCATATTCGTCAGCGCGCGGTCCAACTACT 250
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
67 gGlnLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnTyrT 84

251 CGAGGCGCGAGCAGCAGCAGCAGCGCTGCTCCGAAATGGGCTTC 300
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
84 hrLysThrAspAspGluAlaAsnGlnLeuLeuSerSerLysMetAsnPhe 100
```

seq_name: p1r2:G84693

seq_documentation_block:
probable proline-rich protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: G84693
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: G84693
A;Status: preliminary
A:Molecule type: DNA
A:Residues: 1-891 <STO>
A;Cross-references: GB:AE002093; NID:g3980411; PIDN:AAC95214.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g29210
A;Map position: 2

alignment_scores:
Quality: 119.00 Length: 111
Ratio: 2.052 Gaps: 6
Percent Similarity: 52.252 Percent Identity: 35.135

alignment_block:

US-09-462-480-4/rev x G84693 ..

Align seg 1/1 to: G84693 from: 1 to: 891

```
301 AGAAGCCATTGGGAGGACAGCGCTGCTGCTCTCT...CGTCGGCC 255
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
425 ArgSerProLeuAlaLysArgGlyArgSerAspSerProGlyArgSerPr 441
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
254 CTCGAGTATTGGACGCGGCTGACGACATATTCGTCGAGA..... 215
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
441 oSerProValAlaArgLeuArgAspProThrGlyAlaArgLeuProSerP 458
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
214 .....TCTCGTCGAGTTCTGCTTCTGCTTATTGCTGCTCTTCTTGA 173
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
458 roSerIleGluGlnArgLeuProSerProProValAlaGlnArgLeuPro 474
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
172 AGCGCACACCG...CGGCTGCGGCGGCTCCCGCGCGCGCGCGCCAC 126
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
475 SerProProProArgArgAlaGlyLeuProSerProPro..... 487
125 TGGCCCTGCAACGACCTCGCTGCTCCACCTCCGCTGCTGCTGGTTT 76
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
488 ....ProAlaGlnArgLeuProSerProProProArgArgAlaGlyLeuP 503
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
75 CAGTCGCGCGGAGA.....TCCGCTCGAAATTACCTG..... 44
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
503 roSerProMetArgIleGlySerHisAlaAlaAsnHisLeuGluSer 519
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
43 .....CCTCCTGCCCGA 32
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
520 ProSerProSerSerLeuSerProProGlyArg 530
```

seq_name: p1r2:A40671

seq_documentation_block:

dynamain, internal form 1, long C-terminal form - human

C:Species: Homo sapiens (man)

C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 02-Feb-2001

C;Accession: A40671

R;van der Bliek, A.M.; Redelmeier, T.E.; Damke, H.; Tisdale, E.J.; Meyerowitz, E.M.;

J. Cell Biol. 122, 553-563, 1993

A;Title: Mutations in human dynamain block an intermediate stage in coated vesicle for

A;Reference number: A40671; MUID:93328758

A;Accession: A40671

A;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-864 <VAN>

A;Cross-references: GB:L07807

C;Superfamily: human dynamain II; pleckstrin repeat homology

C;Keywords: alternative splicing; GTP binding; membrane trafficking; nucleotide bindi

F:38-45/Region: nucleotide-binding motif A (P-loop)

F:132-137/Region: nucleotide-binding motif B

F:518-623/Domain: pleckstrin repeat homology <PLK>

alignment_scores:

Quality: 105.50 Length: 96
Ratio: 2.198 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 35.417

alignment_block:

US-09-462-480-4/rev x A40671 ..

Align seg 1/1 to: A40671 from: 1 to: 864

```
301 AGAAGCCATTGGG.....AGGACAGCGCTGCTGCTGC 267
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
773 ArgSerProThrSerSerProThrProGlnArgArgAlaProAlaValPr 789
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
266 TCCTCGTCGGCCCTCGAGTATTGGACGCGCGCTGACGAATATTCTCGCA 217
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
789 oProAlaArgProGlySerArgGlyProAlaProGlyPro..... 802
||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
216 GATCTCGTCGAGTCTCTGCTTATTGCTGCTCTTCTTGGAGCGCA 167
```

```
803 .....ProProAlaGlySer.....AlaLeuGlyAla 812
166 CCACCGCGCTGGCGCGCTCCCGCGCGCGCGCGCGCGCTGC 117
813 ProProValProSerArgProGlyAlaSerProAspProPheGlyProPr 829
116 AACGAACCTGCCGTCG.....ACTCCACCTGGTCGATCTGGGT 79
829 oProGlnValProSerArgProAsnArgAlaProProGlyValProSerA 846
78 TTTCAGCTCCCGGAGATCCGCTCGAAATACCTGCCT 41
846 rgSerGlyGlnAlaSerProSerArgProGluSerPro 858
```

seq_name: pir1:EDBEIF

seq_documentation_block:
immediate-early protein IE180 - suid herpesvirus 1 (strain Indiana-Funkhauser)
C:Species: suid herpesvirus 1
C:Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 16-Feb-1997
C:Accession: S04713
R:Cheung, A.K.

Nucleic Acids Res. 17, 4637-4646, 1989
A:Title: DNA nucleotide sequence analysis of the immediate-early gene of pseudorabies vi
A:Reference number: S04713; MUID:89315207
A:Accession: S04713
A:Molecule type: DNA
A:Residues: 1-1460 <CHE>
C:Superfamily: herpesvirus immediate-early protein IE175
C:Keywords: DNA binding; early protein; transcription regulation

alignment_scores:
Quality: 104.00 Length: 154
Ratio: 1.926 Gaps: 5
Percent Similarity: 35.065 Percent Identity: 25.325

alignment_block:
US-09-462-480-4/rev x EDBEIF ..

Align seg 1/1 to: EDBEIF from: 1 to: 1460

```
298 AGCCATTTCGAGGACAGCGCCTGCTGCTCCTCGTCGCGCCTCGAG 249
115 SerProAlaAlaGlySerProValGlyLeuSerIleArgAlaProSerTh 131
248 TATT.....GGAC 241
131 rValThrSerSerSerglyProGlyProAlaProGlyProGlyA 148
240 GCGGCTGAGCAATATTCGTCGAGATCTGTCGAGTTCCTGCTTCGCT 191
148 rArgProArgGlnHisSerGlnArgGlnArgProGlyProProAlaAla 164
190 TATTGGCTGCTTCTTGAAGCGCACCGCGGCTGGCGCGCTGCCCTCCC 141
165 rGly.....AlaArgProProGlnProProArgProProPr 178
140 GCCG..... 137
178 qProAlaProProAlaProProAlaProProAlaProArgProA 195
137 ..... 137
195 rGlyAspGlyProProArgGlyGlyThrArgSerValSerProGlyArg 211
136 GCGCGCGCACTGGCCTGCAACGAAC..... 110
212 ArgArgGlyLeuGlyProArgArgHisGlnHisSerGlnGlnArgTrpPr 228
109 .....CTGCCGTCGACTCCACCTG 91
```

```
228 oGlnArgArgHisGlyGlyProLeuProGlnProProProProProg 245
90 GTCGATCTGGTTTTCAGGTCCGCGGAGATCCGCTCGAAATACCTGCCT 41
245 lyArgSer.....ArgArgProAlaAlaAlaAlaProPro 256
40 CCTGCCCGAGGG 29
257 ProAlaGluGly 260
```

seq_name: pir2:T35785

seq_documentation_block:
probable beta-glucosidase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
C:Accession: T35785
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998
A:Reference number: Z21570
A:Accession: T35785
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-859 <SEE>
A:Cross-references: EMBL:AL031013; PIDN:CAA19790.1; GSPDB:GN00070; SCOEDB:SC8A6.18
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC8A6.18
C:Superfamily: beta-glucosidase

alignment_scores:
Quality: 99.50 Length: 92
Ratio: 2.117 Gaps: 5
Percent Similarity: 51.087 Percent Identity: 38.043

alignment_block:
US-09-462-480-4 x T35785 ..

Align seg 1/1 to: T35785 from: 1 to: 859

```
39 GGAGGACAGTAAATTCGAGCGGATCTCCGGGACCTGAAACCCAGATCG 88
287 GlyGlyArg.....AspValSerArgSerProValProGlyArg 299
89 ACCAGGT.....GGAGTCGACGCGAGGTTCTGTGCAG 120
299 gProAlaAspSerProGlyAlaGluGlyAlaAspGlyGlyAlaGlyAlaG 316
121 GCCCAGTGGCGCGCGCGGCGGACGCCGCCCGCGCGGT..... 164
316 lyProSerSer.....GlyAlaGluGlyLeuProGlyArgGlyProAla 330
165 ...GGTCGCTTCCAAAGAACGAGCCCAATACGACAGCAGGAACCTCAGC 211
331 HisGlyAlaLysProSerGlyProArgProArgAlaGlyAspGlyArg 347
212 AGATCTCGACGAATATTCGTGACGCCGCGTCCAATACTCGAGCGCGCAG 261
347 gAlaLeu.....AlaArgA 352
262 GAGGACGACGACGAGCGCTGTCCTC 287
352 rGAlaValAlaAlaGlyAlaValLeu 360
```

seq_name: pir2:B75514

seq_documentation_block:
penicillin-binding protein 1 - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: B75514
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J

A: Experimental source: strain A3(2)
C: Genetics:
A: Gene: SC0EDB:SC3A3.06

alignment_scores:

Quality: 96.50 Length: 119
Ratio: 1.856 Gaps: 7
Percent Similarity: 43.697 Percent Identity: 32.773

alignment_block:

US-09-462-480-4/rev x T34871 ..

Align seg 1/1 to: T34871 from: 1 to: 509

```
301 AGAAGCCATTGTCGAGGACAGCGCTGCTGCTCTCTCTCGG..... 257
|||||  |||  ::::|  |||  |||||  |||
397 ArgSerProSerAlaProAlaAlaHisAlaProAlaProAspArgSerPr 413
:::  :::  |||||  |||||  |||||  |||
256 .CCCTCGAGTATTGAGCG.....CGCCCTGACGAATATTCGTCGAGAT 214
||||  |||||  |||||  |||||  :::  |||
413 oProAspProAspGlyArgProSerAspProAspArgSerSerProAla 430
|||  |||||  |||||  |||||  |||||  |||
213 CTCGTCGAGTCCCTCTCTCT..... 191
|||  |||||  |||||  |||||  |||||  |||
430 erAspArgLeuProProAlaSerAspArgSerProSerAlaProAlaAla 446
|||  |||||  |||||  |||||  |||||  |||
190 { .....TATTGCTGCTCTTGGAGCGCACCACCGCGGCC 156
|||  |||||  |||||  |||||  |||||  |||
447 HisAlaProAlaProAspTrp.....SerProProAspPr 458
|||  |||||  |||||  |||||  |||||  |||
155 T...GGGGGGCGTCCCGCGCGCGCGGCACCTGGCCCTGCAACGAAC 109
|||  |||||  |||||  |||||  |||||  |||
458 oAspGlyArgProSerAspProAspArgSerSerProAlaSerAspArgL 475
|||  |||||  |||||  |||||  |||||  |||
108 TGGCTGCACTCACCTGGTCGATCTGGTTTTCAGGTGCGCGGAGATCC 59
||||  |||||  |||||  |||||  |||||  |||
475 euProProAlaPro..... 479
58 GCTCGAATATACCTGCTCTGCTCCGAGGTAGCGCATCGGTCTTCATC 9
:::  |||||  |||||  |||||  |||||  |||
480 .....AspArgLeuProProAla.....ProAspArgSerProSe 491
8 TCTGCCA 2
|  |||
491 rAlaPro 493
```

seq_name: pir2:H72706

seq_documentation_block:

hypothetical protein APE1071 - Aeropyrum pernix (strain K1)
C: Species: Aeropyrum pernix
C: Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C: Accession: H72706
R: Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudooh, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999
A: Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A: Reference number: A72450; MUID: 99310339
A: Accession: H72706
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-117 <KAW>
A: Cross-references: DDBJ:AP000060; NID: g5104188; PIDN: BAA80056.1; PID: d1043842; PID: g5104188
C: Genetics:
A: Gene: APE1071

alignment_scores:

Quality: 95.00 Length: 110
Ratio: 1.827 Gaps: 10
Percent Similarity: 47.273 Percent Identity: 39.091

alignment_block:

US-09-462-480-4/rev x H72706 ..

Align seg 1/1 to: H72706 from: 1 to: 117

```
298 AGCCATTGTCGAGGACAGCGCTGCTGCTCTCTCTCTCGGCCCTCGAG 249
|||||  |||||  |||  :::  |||  |||
11 SerProGlyAlaArgProGluProProSerHisTyrArgThrPro..... 25
248 TATTGAGCGCGG...CCTGACGAATATTCGTCGAGA...TCTCGTCGAG 205
|||||  |||||  |||  :::  |||||  |||||
26 ....GlyArgArgGlyProArgCysHisProSerArgGlySerArgArgG 41
204 TTCCT.....GCTCTCTTATTGCTGCTCTTCTTGGAGGCGCACCA 164
|||  |||  :::  |||  |||  :::
41 lnProLeuTrpCysArgThrGlnAlaTrpProTrpArgGlyArgSerPro 57
163 CCG.....CGCGCTGGCGG.....CCGCTC 144
|||  |||||  |||||  |||  :::
58 ProCysLeuLeuCysProGlyArgGluProSerSerArgTyrProProGl 74
143 CCCGCGCGCGCGCCACT.....GGCCCTGCAACGAACCTGCGCTCG. 101
:::  |||||  :::  |||||  |||||  |||
74 uproProSerHisThrThrProHisGlyProArgGlylleGlnProSerL 91
100 .....ACTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 59
|||||  |||||  |||||  |||||  |||||  |||
91 euGlnProProProGlyArgTyr.....ArgArgArgSerGln 103
58 .....GCTCGAATATACCTGCTCTCT 38
|||  |||||  |||||  |||||
104 ProProArgValSerSerTyrProProPro 113
```

seq_name: pir1:QB072

seq_documentation_block:

microtubule-associated protein tau, form 3 - bovine
N: Contains: microtubule-associated protein tau, form 4; microtubule-associated protein tau, form 3
C: Species: Bos primigenius taurus (cattle)
C: Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 31-Mar-1996
R: Himmler, A.; Drechsel, D.; Kirschner, M.W.; Martin Jr., D.W. Mol. Cell. Biol. 9, 1381-1388, 1989
A: Title: Tau consists of a set of proteins with repeated C-terminal microtubule-binding motifs
A: Reference number: A31939; MUID: 89261765
A: Accession: B31939
A: Molecule type: mRNA
A: Residues: 1-402 <HIM>
A: Cross-references: GB:M26157; GB:M26158
R: Paudel, H.K.; Lew, J.; Ali, Z.; Wang, J.H. J. Biol. Chem. 268, 23512-23518, 1993
A: Title: Brain proline-directed protein kinase phosphorylates tau on sites that are a part of the microtubule-binding domain
A: Reference number: A48885; MUID: 94043150
A: Accession: A48885
A: Molecule type: protein
A: Residues: 'X', 157-162, 'X', 164-165, 'X', 167-170; 192-195, 'X', 197-201, 'X', 358-364, 'X', 3
A: Experimental source: brain
A: Note: sequence modified after extraction from NCBI backbone
R: Aizawa, H.; Kawasaki, H.; Murofushi, H.; Kotani, S.; Suzuki, K.; Sakai, H. J. Biol. Chem. 263, 7703-7707, 1988
A: Title: Microtubule-binding domain of Tau proteins.
A: Reference number: A28173; MUID: 88227970
A: Accession: A28173
A: Molecule type: protein
A: Residues: 159-172, 'X', 174-177 <AIZ>
C: Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
C: Keywords: alternative splicing; microtubule binding; phosphoprotein; tandem repeat
F: 1-402/Product: microtubule-associated protein tau, form 3 #status predicted <BT4>
F: 1-234, 297-402/Product: microtubule-associated protein tau, form 5 #status predicted <BT4>
F: 101-402/Product: microtubule-associated protein tau, form 4 #status predicted <BT7>
F: 159-177/Region: microtubule binding #status experimental

OM of: US-09-462-480-4 to: SwissProt_40:* out_format : pfs
Date: Jul 22, 2002 1:41 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet_n2p_model -DEV=xlh
-O=/cnr2.1/USPTO_spool/US09462480/runat_18072002_164420_19600/app_query.fasta_1.2850
-DB=SwissProt_40 -QPM=fastan -SUFFIX=rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-XGAPOP=4.000 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM_ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480.ecgn1.1.29 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-4

Query length: 302

Database: SwissProt_40:*

Database sequences: 105224

Database length: 38719550

Search time (sec): 49.760000

score_list:

Sequence	Strd	Orig	zScore	Escore	Len
SwissProt_40:CF10_MYCTU	+	481.00	710.03	7.6e-32	99
SwissProt_40:CF10_MYCLE	+	182.00	274.67	1.3e-07	99
SwissProt_40:DYN1_HUMAN	+	105.50	146.80	0.2040	864
SwissProt_40:YX2_CAEEL	+	97.50	145.99	0.9410	208
SwissProt_40:DYN1_RAT	-	97.50	135.27	0.9092	851
SwissProt_40:DYN1_MOUSE	-	97.50	135.18	0.9089	861
SwissProt_40:Y135_STRFR	-	93.50	136.25	1.96	348
SwissProt_40:Y135_PVKA	-	93.50	125.41	1.89	1446
SwissProt_40:CUT1_CANFA	-	93.00	127.68	2.10	975
SwissProt_40:CBPA_DICDI	-	92.00	131.83	2.58	467
SwissProt_40:WAIP_HUMAN	-	91.50	130.53	2.82	503
SwissProt_40:TAU_MOUSE	-	91.50	127.68	2.80	732
SwissProt_40:CC19_CAEEL	-	90.50	133.45	3.45	283
SwissProt_40:WAS3_HUMAN	-	90.00	128.36	3.74	502
SwissProt_40:NNE3_RAT	-	90.00	121.50	3.66	1237
SwissProt_40:GSRI_HUMAN	-	90.00	119.99	3.64	1509
SwissProt_40:ROK_HUMAN	-	89.50	128.25	4.11	463
SwissProt_40:ROK_RABIT	-	89.50	128.25	4.11	463
SwissProt_40:ROK_MOUSE	-	89.50	128.24	4.11	464
SwissProt_40:WASP_HUMAN	-	89.50	127.64	4.10	502
SwissProt_40:ES6D_MYCTU	+	89.00	138.96	4.68	103
SwissProt_40:IRS2_HUMAN	-	89.00	119.53	4.40	1324
SwissProt_40:WS14_MOUSE	-	88.50	122.05	4.88	864
SwissProt_40:IF2_MYCLE	+	88.00	120.81	5.35	924
SwissProt_40:ANTI1_ONCVO	+	87.00	113.40	6.33	2022
SwissProt_40:S3A2_MOUSE	-	86.00	123.14	7.90	464
SwissProt_40:RP3A_BOVIN	-	85.50	119.24	8.59	704
SwissProt_40:IF2_MYCTU	+	85.50	117.37	8.54	900
SwissProt_40:VGLB_HSVBP	-	85.50	117.14	8.53	928
SwissProt_40:VGLB_HSVBC	-	85.50	117.11	8.53	932
SwissProt_40:SFR2_CHICK	+	85.00	127.33	9.70	221
SwissProt_40:ICP3_HSVF	+	85.00	126.00	9.66	263
SwissProt_40:SYN1_HUMAN	-	84.50	117.77	10.35	705
SwissProt_40:TRX2_HUMAN	-	84.50	107.52	10.02	2715
SwissProt_40:YML2_THFE	-	84.00	131.32	11.90	108
SwissProt_40:ICP3_HSVF	-	84.00	124.87	11.65	252
SwissProt_40:NIFI_MOUSE	-	84.00	122.42	11.56	348
SwissProt_40:CCPA_ACEXY	-	84.00	122.31	11.56	353
SwissProt_40:T2D3_HUMAN	-	84.00	113.78	11.25	1083

SwissProt_40:TAU_RAT - 83.50 115.84 12.46 751 ! P19332 rattus norvegicus (rat)
SwissProt_40:MYSD_DICDI - 83.00 112.12 13.55 1113 ! P34109 dictyostellium discoidi
SwissProt_40:ICPA_HVMG - 83.00 110.29 13.47 1415 ! Q02362 marek's disease 1
SwissProt_40:VNUA_PVKA - 83.00 108.75 13.40 1733 ! P33485 pseudorabies virus
SwissProt_40:SCE3_SCHPO - 82.50 119.40 15.26 388 ! O14369 schizosaccharomyces

seq_name: SwissProt_40:CF10_MYCTU

seq_documentation_block:

ID CF10_MYCTU STANDARD; PRT; 99 AA.
AC Q69739;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 10 kDa culture filtrate antigen cfp10.
GN CFP10 OR LHP OR RV3874 OR MT3988 OR MTV027.09.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
RC STRAIN=H37RV;
RX MEDLINE=99061212; PubMed=9846755;
RA Berthet F.X., Rasmussen P.B., Rosenkrands I., Andersen P.,
RA Gicquel B.;
RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel
RT low-molecular-mass culture filtrate protein (CFP-10).";
RL Microbiology 144:3195-3203(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF004671; AAC83445.1; -
CC EMBL: AL022120; CAA17966.1; -
CC EMBL: AE007190; AAK48356.1; -
CC TIGR: MT3988; -
CC TuberculList: RV3874; -
CC Complete proteome.
CC INIT_MET 0
CC SEQUENCE 99 AA; 10663 MW; EBCA65A996C5489D CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Y14967; CAA75210.1;
 DR EMBL; AL583917; CAC29558.1;
 DR Leproma; ML0050;
 KW Complete proteome.
 FT INIT_MET 0 0 BY SIMILARITY
 SQ SEQUENCE 99 AA; 10833 MW; 859B484F7EFE5A8A CRC64;

alignment_scores:
 Quality: 182.00 Length: 99
 Ratio: 2.333 Gaps: 0
 Percent Similarity: 78.788 Percent Identity: 39.394

alignment_block:
 US-09-462-480-4 x CF10_MYCLE

Align seg 1/1 to: CF10_MYCLE from: 1 to: 99

4 CGACGAGTGAAGACCGATCCCGCTACCTCGGGCAGGAGGAGGTAATTT 53
 1 AlaGluMetIleThrGluAlaAlaIleLeuThrGlnGlnAlaAlaGlnPh 17
 54 CGAGCGGATCTCCGGCGACCTGAAACCCAGATCCAGGTTGGAGTGA 103
 17 eAspGlnIleAlaSerGlyLeuSerGlnGluArgAsnPheValAspSerI 34
 104 CGGCGAGTTGTCGAGCGCCAGTCGCGCGCGGGGAGCGGCGCCGCC 153
 34 leglyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAlaAla 50
 154 CAGCGCGCGTGGTTCCTCCAGAACGACGACCAATAAGCAGGAGCAGGA 203
 51 LeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleArgG 67
 204 ACTCGACGAGATCTCGACGAAATATTCGTCAGGCGCGGCTCCCAATCT 253
 67 nLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnThrL 84
 254 GGGCGAGCAGGAGCAGCAGGCGGCTGCTCTCGCAATGGGCTTC 300
 84 yThrAspAspGluAlaAsnGlnLeuLeuSerSerLysMetAsnPhe 99

seq_name: SwissProt_40:DYN1_HUMAN

seq_documentation_block:

ID DYN1_HUMAN STANDARD; PRT; 864 AA.
 AC Q05153;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dynamin-1 (EC 3.6.1.50).
 GN DNM1 OR DNM
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93328756; PubMed=8101525;
 RA van der Bliek A.M., Redelmeier T.E., Tisdale E.J., Meyerowitz E.M.,
 RA Schmid S.L.;
 RT "Mutations in human dynamin block an intermediate stage in coated
 RT vesicle formation."
 RL J. Cell Biol. 122:553-563(1993).
 RN [2]
 RP STRUCTURE BY NMR OF 511-630.
 RX MEDLINE=95153276; PubMed=7850421;

alignment_scores:
 Quality: 481.00 Length: 99
 Ratio: 4.859 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.990

alignment_block:
 US-09-462-480-4 x CF10_MYCTU

Align seg 1/1 to: CF10_MYCTU from: 1 to: 99

4 CGACGAGTGAAGACCGATCCCGCTACCTCGGGCAGGAGGAGGTAATTT 53
 1 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 17
 54 CGAGCGGATCTCCGGCGACCTGAAACCCAGATCCAGGTTGGAGTGA 103
 17 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerT 34
 104 CGGCGAGTTGTCGAGCGCCAGTCGCGCGCGGGGAGCGGCGCCGCC 153
 34 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 50
 154 CAGCGCGCGTGGTTCCTCCAGAACGACGACCAATAAGCAGGAGCAGGA 203
 51 GlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 67
 204 ACTCGACGAGATCTCGACGAAATATTCGTCAGGCGCGGCTCCCAATCT 253
 67 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTrpSer 84
 254 GGGCGAGCAGGAGCAGCAGGCGGCTGCTCTCGCAATGGGCTTC 300
 84 rgAlaAspGluGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 99

seq_name: SwissProt_40:CF10_MYCLE

seq_documentation_block:

ID CF10_MYCLE STANDARD; PRT; 99 AA.
 AC Q33084;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 10 kDa culture filtrate antigen cfpl0 homolog.
 GN M0050 OR MLCB628.13C.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Eiglmeier K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;
 RA Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation

Downing A.K., Driscoll P.C., Gout I., Salim K., Zvelebil M.J.,
Waterfield M.D.;
"Three-dimensional solution structure of the pleckstrin homology
domain from dynamin.";
Curr. Biol. 4:884-891(1994).
[3]
X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 509-633.
MEDLINE=95042718; PubMed=7954789;
Ferguson K.M., Lemmon M.A., Schlessinger J., Sigler P.B.;
"Crystal structure at 2.2-A resolution of the pleckstrin homology
domain from human dynamin.";
Cell 79:199-209(1994).
[4]
X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 509-630.
MEDLINE=95360729; PubMed=7634088;
Timm D., Salim K., Gout I., Guruprasad L., Waterfield M., Blundell T.;
"Crystal structure of the pleckstrin homology domain from dynamin.";
Nat. Struct. Biol. 1:782-788(1994).
CC -!- FUNCTION: MICROTUBULE-ASSOCIATED FORCE-PRODUCING PROTEIN INVOLVED
CC IN PRODUCING MICROTUBULE BUNDLES AND ABLE TO BIND AND HYDROLYZE
CC GTP. MOST PROBABLY INVOLVED IN VESICULAR TRAFFICKING PROCESSES, IN
CC PARTICULAR ENDOCYTOSIS.
CC -!- CATALYTIC ACTIVITY: GTP + H(2O) = GDP + phosphate.
CC -!- SUBCELLULAR LOCATION: MICROTUBULE-ASSOCIATED.
CC -!- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L07807; AAA02803.1; -;
CC EMBL; L07808; AAA02804.1; ALT_SEQ.
CC EMBL; L07809; AAA02805.1; ALT_SEQ.
CC EMBL; L07810; AAA02806.1; ALT_SEQ.
CC PDB; 1DXN; 27-FEB-95.
CC PDB; 2DXN; 12-NOV-97.
CC MIM; 602377; -;
CC InterPro; IPR001401; Dynamin.
CC InterPro; IPR000375; Dynamin_central.
CC InterPro; IPR003130; GED.
CC InterPro; IPR001849; PH.
CC Pfam; PF00350; dynamin; 1.
CC Pfam; PF01031; dynamin_2; 1.
CC Pfam; PF02212; GED; 1.
CC Pfam; PF00169; PH; 1.
CC PRINTS; PR00195; DYNAMIN.
CC SMART; SM00553; DYNC; 1.
CC SMART; SM00302; GED; 1.
CC SMART; SM00233; PH; 1.
CC PROSITE; PS00410; DYNAMIN; 1.
CC PROSITE; PS50003; PH_DOMAIN; 1.
CC Hydrolase; Motor protein; GTP-binding; Microtubules; Multigene family;
KW Endocytosis; 3D-structure.
FT NP_BIND 38 45 GTP (BY SIMILARITY).
FT NP_BIND 136 140 GTP (BY SIMILARITY).
FT NP_BIND 205 208 GTP (BY SIMILARITY).
FT DOMAIN 519 625 PH.
SQ SEQUENCE 864 AA; 97407 MW; 7B2D86B572FFE0EF CRC64;

alignment_scores:
Quality: 105.50 Length: 96
Ratio: 2.198 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 35.417

alignment_block:
US-09-462-480-4/rev x DYN1_HUMAN

Align seg 1/1 to: DYN1_HUMAN from: 1 to: 864

301 AGAAGCCCATTTGCG.....AGGACAGCGCTGCTGCTGC 267
|||||
773 ArgSerProThrSerSerProThrProGlnArgAlaProAlaValPr 789
266 TCCTCGTCGCGCCCTCGAGTATTTGGACGCCGCGCTGACGAATATTTCGTCGA 217
||| |||||
789 oProAlaArgProGlySerArgGlyProAlaProGlyPro..... 802
216 GATCTCGTCGAGTTCCTGCTCTCTATTGCTGCTTCTTTGGAGCGCA 167
|||||
803ProProAlaGlySer.....AlaLeuGlyGlyAla 812
166 CCACCGCGGCGCTGCGCGGCTGCTGCCGCGCGCGGCGGCGGCGGCGCTGCG 117
|||||
813 ProProValProSerArgProGlyAlaSerProAspPropheGlyProPr 829
116 AACGAACCTGCGCTG.....ACTCCACCTGCTGCTGATCGGCT 79
: : : : :
829 oProGlnValProSerArgProAsnArgAlaProProGlyValProSerA 846
78 TTTTCAGTCGCGCGGAGATCCGCTCGAAATTACCTGCCT 41
|||||
846 rgSerGlyGlnAlaSerProSerArgProGluSerPro 858

seq_name: SwissProt_40:YSX2_CAEEL

seq_documentation_block:
ID YSX2_CAEEL STANDARD; PRT; 208 AA.
AC Q10021;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 24.0 kDa protein T28D9.2 in chromosome II.
GN T28D9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U28738; AAA68314.1; -;
CC WormPep; T28D9.2; CE02067.
CC InterPro; IPR000504; RRM.
CC Pfam; PF00076; rrm; 1.
CC SMART; SM00360; RRM; 1.
CC PROSITE; PS50102; RRM; 1.
CC PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW Hypothetical protein; RNA-binding.
FT DOMAIN 2 74 RNA-BINDING (RRM).
FT DOMAIN 84 205 ARG/SER-RICH.
SQ SEQUENCE 208 AA; 23992 MW; 0FE8AAD05FB777EF CRC64;

alignment_scores:
Quality: 97.50 Length: 70
Ratio: 2.267 Gaps: 3
Percent Similarity: 61.429 Percent Identity: 42.857

CC	-1- SIMILARITY: BELONGS TO THE CUT FAMILY OF HOMEBOX PROTEINS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC	or send an email to license@isb-sib.ch).
CC	-----
CC	EMBL X69017: CAA48782.1; -
DR	DR PIR: S33121; S33121.
DR	HSP: P10037; IAU7.
DR	TRANSFAC: T01485; -
DR	TRANSFAC: T01995; -
DR	InterPro: IPR003350; CUT.
DR	InterPro: IPR001356; Homeobox.
DR	Pfam: PF02376; CUT; 3.
DR	Pfam: PF00046; homeobox; 1.
DR	SMART: SM00389; HOX; 1.
DR	PROSITE: PS00027; HOMEBOX_1; 1.
DR	PROSITE: PS00071; HOMEBOX_2; 1.
KW	Transcription regulation; Homeobox; DNA-binding;
KW	Developmental protein; Nuclear protein; Repeat; Repressor;
KW	Coiled coil; Alternative splicing.
FT	NON_TER 1
FT	DNA_BIND <1 67 CUT 1.
FT	DOMAIN 113 169 COILED COIL (POTENTIAL).
FT	DNA_BIND 374 461 CUT 2.
FT	DNA_BIND 557 641 CUT 3.
FT	DNA_BIND 684 743 CUT 4.
FT	SEQUENCE 975 AA; I05428 MW; 724243B32C5BCFD9 CRC64;
CC	-----

```

alignment_scores:
    Quality: 93.00      Length: 125
    Ratio:   1.550     Gaps:    8
Percent Similarity: 48.000 Percent Identity: 32.800

alignment_block:
US-09-462-480-4/rev x CUTL_CANFA ..

Align seg 1/1 to: CUTL_CANFA from: 1 to: 975

283 ACAGCGCTGCTGCTGCTCTCTCGTCGG.....CCCTCGAGTATTGG 243
       |||:|||||:|||||:|||:::||||:::  ::
795 SerAlaProAlaAlaAlaAlaSerGlnGlyGlyProAlaGluAlaI 811
       |||:|||||:|||||:|||:::||||:::  ::
242 ACGCGCGCCTGCACGAATATTTCGTGCAGATCTCTCGAGTTCTGCTCTG 193
       |||:|||||:|||||:|||:::||||:::  ::
811 aValaLaProGluGluArgGluGluAlaProArg.....ProAlaGluL 826
       |||:|||||:|||||:|||:::||||:::  ::
192 CTTATTGGCTCTTCTTGGAAGGCCACCACCGCGGCTGGGCGGCGCTGC 143
       |||:|||||:|||||:|||:::||||:::  ::
826 ys.....ArgSerArgargProArgGlyProGlyProGly 837
       |||:|||||:|||||:|||:::||||:::  ::
142 CGCGCGCGCGCG...GCCACTGGCCCTGCAACCAACTGCCGTCGACTCC 96
       |||:|||||:|||||:|||:::||||:::  ::
838 ProGlyArgArgGlyGlyGlyGlyProAlaProGlyAlaProAlaIaPr 854
       |||:|||||:|||||:|||:::||||:::  ::
95 A.....CCTGGTCGATCTGGGCTTTT 76
       |||:|||||:|||||:|||:::||||:::  ::
854 oAlaAlaAlaAlaArgGlyProSerArgArgProGlyAlaArgAlaLysP 871
       |||:|||||:|||||:|||:::||||:::  ::
75 CAGTCTCGCGGAGA.....TCGCTCGA 53
       |||:|||||:|||||:|||:::||||:::  ::
871 roArgArgArgArgArgArgHisAlaArgGlyGlyGlyArg 887
       |||:|||||:|||||:|||:::||||:::  ::
52 AATTACTCTGCT..CCTGCCCGAGGG.....TA 27
       |||:|||||:|||||:|||:::||||:::  ::
888 ArgTyrlLeuSerArgProAlaArgGlyGlyProCysAlaArgAspGI 904
       |||:|||||:|||||:|||:::||||:::  ::
26 GCGGCATCGGTTTCATCTCTCTGCCA 2

```

```

|||||
904 yAlaHisArgSerAlaLeuPro 912

seq_name: SwissProt_40:CBPA_DICDI
seq_documentation_block:
ID ID CBPA_DICDI STANDARD; PRT; 467 AA.
AC P35085;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Calcium-binding protein.
DE CBPA.
GN Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RS SEQUENCE FROM N.A.
RP RP
RC STRAIN=AX2;
RA Wennington R., Greenwood M., Tsang A.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U03413; AAA03471.1; -.
DR Dictydb; DD01055; cbpa.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 2.
DR SMART; SM00054; EFh; 2.
DR PROSITE; PS00018; EF_HAND; 1.
DR Calcium-binding.
KW
FT CA_BIND 412 423 EF-HAND (POTENTIAL).
SQ SEQUENCE 467 AA; 489976 MW; 2D2055D1E344FC18 CRC64;

```

```

alignment_scores:          Quality: 92.00      Length: 91
                          Ratio: 1.917      Gaps: 3
Percent Similarity: 52.747  Percent Identity: 26.374

alignment_block:
US-09-462-480-4/rev x CBPA_DICDI ..

Align seg 1/1 to: CBPA_DICDI from: 1 to: 467

295 CCATTCGTGAGACAGCGCTGCTGCTGCTCGTCGGCCCTCGAGTAT 246
|||||: ||| :|||:||||| |||:|||||
60 ProTyrProGlyThrGlnGlnProGlyAlaProGlyAlaProGlyGlnTy 76
|||||: ||| :|||:||||| |||:|||||
245 T...GGACGCCGCCCTGACGAATATTCGTGCAGATCTCGTCAGTTCCTG 199
|||||: ||| :|||:||||| |||:|||||
76 rProProGlnGlnProGlyGlnTyrProProGlnGlnProGlyAlaPro. 92
|||||: ||| :|||:||||| |||:|||||
198 CTTCTGCTTATGGCTGCTTCTTGGAAAGCCACACCG...CGGCCTGGG 152
|||||: ||| :|||:||||| |||:|||||
93 .....GlyGlnTyrProProGlnGlnProGly 101
|||||: ||| :|||:||||| |||:|||||
151 CGGCGCTCCCGCGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 102
|||||: ||| :|||:||||| |||:|||||
102 GlnProGlyTyrProProGlnGlnProGlyGlnSerGlyGlnTyrProPr 118
|||||: ||| :|||:||||| |||:|||||
101 GACTCCACCTGTCGATCTGGTTCAGTTCGCGCGGAGATCCGCTCGAA 52
|||||: ||| :|||:||||| |||:|||||
118 oGlnGlnProGlyGlnProGlyTyrProProGlnGlnProGlyAlaPro 135
|||||: ||| :|||:||||| |||:|||||
51 ATTACCTGCTCTCTGCGCCGAGG 29

```

```

:: ||||| ::|||
135 lyGlnTyProProGlnGly 142
seq_name: SwissProt_40:WAIP_HUMAN

seq_documentation_block:
ID WAIP_HUMAN STANDARD; PRT; 503 AA.
AC Q43516; Q9UNP1; Q15220;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Wiskott-Aldrich syndrome protein interacting protein (WASP interacting
DE protein) (PRPL-2 protein).
GN WASPIP OR WIP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=98070810; PubMed=9405671;
RA Ramesh N., Anton I.M., Hartwig J.H., Geha R.S.;
RT "WIP, a protein associated with Wiskott-Aldrich syndrome protein,
RT induces actin polymerization and redistribution in lymphoid cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:14671-14676(1997).
RN [2]
SEQUENCE FROM N.A.
RX TISSUE=Tonsil;
RA Kreideweiss S., Delany-Heiken P., Nordheim A., Ruhlmann A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF 192-503 FROM N.A., AND VARIANT ALA-495.
RX MEDLINE=99218549; PubMed=10202051;
RA Stewart D.M., Tian L., Nelson D.L.;
RT "Mutations that cause the Wiskott-Aldrich syndrome impair the
RT interaction of Wiskott-Aldrich syndrome protein (WASP) with WASP
RT interacting protein.";
RL J. Immunol. 162:5019-5024(1999).
CC -|- FUNCTION: MAY HAVE DIRECT ACTIVITY ON THE ACTIN CYTOSKELETON.
CC -|- INDUCES ACTIN POLYMERIZATION AND REDISTRIBUTION.
CC -|- SUBUNIT: BINDS TO WASP, PROFILIN AND ACTIN.
CC -|- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PERIPHERAL BLOOD
CC MONONUCLEAR CELLS, SPLEEN, PLACENTA, SMALL INTESTIN, COLON,
CC THYMUS. LOWER EXPRESSION IN OVARY, HEART, BRAIN, LUNG, LIVER,
CC SKELETAL MUSCLE, KIDNEY, PANCREAS, PROSTATE, TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF031588; AAC03767.1;
DR EMBL; AF106062; AAD45972.1;
DR EMBL; X86019; CAA60014.1;
DR MIM; 602357;
DR InterPro; IPR003124; WH2.
DR Pfam; PF02205; WH2; 1.
DR SMART; SM00246; WH2; 1.
KW Actin-binding; Repeat.
FT DOMAIN 2 13
FT DOMAIN 64 96
FT DOMAIN 241 244
FT DOMAIN 264 433
FT REPEAT 352 361
FT REPEAT 374 383
FT REPEAT 410 419
FT SITE 45 48
FT VARIANT 495 495
FT G -> A.
FT /FTID=VAR_010295.
FT PHRPLR -> SSQAPPP (IN REF. 3).
FT CONFLICT 303 309

FT CONFLICT 360 360 P -> PV (IN REF. 2).
FT CONFLICT 487 503 GSNRRRGPPPLPIPR -> EYFCQGF (IN REF. 2).
SQ SEQUENCE 503 AA; 51489 MW; 43EB8674DD3BF1A CRC64;

alignment_scores:
Quality: 91.50 Length: 107
Ratio: 1.906 Gaps: 5
Percent Similarity: 44.860 Percent Identity: 33.645

alignment_block:
US-09-462-480-4/rev x WAIP_HUMAN
Align seg 1/1 to: WAIP_HUMAN from: 1 to: 503
277 CCTGCTGCTGCTCTCGTGGCGCCCTCGAGTATTGGA..... 242
||||| ||||| ||||| ||||| ||||| ||||| |||||
316 ProGlyProProLeuProLeuProSerSerSerGlyAsnAspGluThrPr 332
241 .CGCCGGCCTGACGAATATTCGTCGAGATCTCGTGGCTCTGCTGCTG 193
||| |||:||||: ||| ||| ||| |||
332 oArgLeuProGlnArgAsnLeuSerLeuSerSerThrProLeuP 349
192 CTTATTGGCTCTTCTTGGAGCCGACCACCGCGCGCTGGG....CGGCCG 146
||| ||| ||||| ||||| ||||| |||||
349 roSerProGlyArgSerGlyProLeuProProProSerGluArgPro 365
145 TCCCGCGCGCGCGC.....GCCACTGGCCCT..... 119
||||| ||| :|||||
366 ProProValArgAspProProGlyArgSerGlyProLeuProProPr 382
118 .....GCAACGAACCTGCCGTCGACTC 97
382 oProProValSerArgAsnGlySerThrSerArgAlaLeuProAlaThrP 399
96 CA.....CCTGGTCGATCGGTTTCAGTCCCGGAGATCCGCTCGA 53
|| ||||| ||||| ||||| ||||| |||||
399 roGlnLeuProSerArgSerGlyValAspSerProArgSerGlyProArg 415
52 AATTACCTGCCTCGCCCGA 32
||||| |||
416 ProProLeuProProAspArg 422

seq_name: SwissProt_40:TAU_MOUSE

seq_documentation_block:
ID TAU_MOUSE STANDARD; PRT; 732 AA.
AC P10637; P10638; Q60684; Q60685; Q60686; Q62286;
DT 01-JUL-1989 (Rel. 11, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Microtubule-associated protein tau (Neurofibrillary tangle protein)
DE (Paired helical filament-tau) (PHF-tau).
GN MAPT OR MTAPT OR TAU.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A. (ISOFORM PNS-TAU).
RX TISSUE=Neuroblastoma;
RX MEDLINE=92262443; PubMed=1374898;
RA Couchie D., Mavilia C., Georgieff I.S., Liem R.K.H., Shelanski M.L.,
RA Nunez J.;
RT "Primary structure of high molecular weight tau present in the
RT peripheral nervous system";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4378-4381(1992).
RN [2]
SEQUENCE FROM N.A. (ISOFORMS TAU-A; TAU-D AND TAU-E).
RX STRAIN=Him OF1; TISSUE=Liver, Kidney, and Brain;
RX MEDLINE=95012085; PubMed=7927211;
RA Kenner L., el-Shabrawi Y., Hutter H., Forstner M., Zatloukal K.,
RA Hoefler G., Preisegger K.-H., Kurzbaue R., Denk H.;
```

RT "Expression of three- and four-repeat tau isoforms in mouse liver.";
 RL Hepatology 20:1086-1089(1994).
 RN [3]
 RX SEQUENCE FROM N.A. (ISOFORMS TAU-B AND TAU-C).
 RC TISSUE=Brain;
 RX MEDLINE=88099510; PubMed=3122323;
 RA Lee G., Cowan N.J., Kirschner M.;
 RT "The primary structure and heterogeneity of tau protein from mouse
 brain.";
 RL Science 239:285-288(1988).
 RN [4]
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM B).
 RC STRAIN=ICR; TISSUE=Brain;
 RX MEDLINE=95182802; PubMed=7877441;
 RA Sawa A., Oyama F., Matsushita M., Ihara Y.;
 RT "Molecular diversity at the carboxyl terminus of human and rat tau.";
 RL Brain Res. Mol. Brain Res. 27:111-117(1994).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=94005827; PubMed=8402267;
 RA Couchie D., Gache Y., Mavillia C., Guilleminot J., Bridoux A.-M.,
 RA Nivez M.-P., Nunez J.;
 RT "High molecular weight tau proteins and acquisition of neuronal
 polarity in peripheral nervous system.";
 RL C. R. Acad. Sci., III, Sci. vie 316:404-409(1993).
 CC -!- FUNCTION: PROMOTES MICROTUBULE ASSEMBLY AND STABILITY, AND MIGHT
 CC BE INVOLVED IN THE ESTABLISHMENT AND MAINTENANCE OF NEURONAL
 CC POLARITY. THE C-TERMINUS BINDS AXONAL MICROTUBULES WHILE THE N-
 CC TERMINUS BINDS NEURAL PLASMA MEMBRANE COMPONENTS, SUGGESTING THAT
 CC TAU FUNCTIONS AS A LINER PROTEIN BETWEEN BOTH. AXONAL POLARITY IS
 CC DETERMINED BY TAU LOCALIZATION (IN THE NEURONAL CELL) IN THE
 CC DOMAIN OF THE CELL BODY DEFINED BY THE CENTROSOME. THE SHORT
 CC ISOFORMS ALLOW PLASTICITY OF THE CYTOSKELETON WHEREAS THE LONGER
 CC ISOFORMS MAY PREFERENTIALLY PLAY A ROLE IN ITS STABILIZATION.
 CC -!- SUBCELLULAR LOCATION: MOSTLY FOUND IN THE AXONS OF NEURONS, IN THE
 CC CYTOSOL AND IN ASSOCIATION WITH PLASMA MEMBRANE COMPONENTS.
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; PNS-TAU (SHOWN HERE),
 CC TAU-A, TAU-B, TAU-D AND TAU-E; ARE PRODUCED BY ALTERNATIVE
 CC SPLICING. THEY DIFFER FROM EACH OTHER BY THE PRESENCE OR ABSENCE
 CC OF UP TO 5 OF THE 14 EXONS. ONE OF THESE OPTIONAL EXONS CONTAINS
 CC THE ADDITIONAL TAU/MAP REPEAT. TWO DIFFERENT C-TERMINI ARE
 CC OBTAINED EITHER BY THE RETENTION OR THE SPLICING OF INTRON 13/14.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN NEURONS AND AT A LOWER LEVEL IN
 CC THE LIVER AND KIDNEY. PNS-TAU IS EXPRESSED IN THE PERIPHERAL
 CC NERVOUS SYSTEM WHILE THE OTHERS ARE EXPRESSED IN THE CENTRAL
 CC NERVOUS SYSTEM.
 CC -!- DEVELOPMENTAL STAGE: SHORTER FORMS OR LOW MOLECULAR WEIGHT TAU
 CC (LMW-TAU) ARE GENERALLY EXPRESSED AT EARLY DEVELOPMENT STAGES AND
 CC LONGER FORMS OR HIGH MOLECULAR WEIGHT TAU (HMW-TAU) IN THE ADULT
 CC BRAIN.
 CC -!- DOMAIN: THE TAU/MAP REPEAT BINDS TO TUBULIN. TYPE I ISOFORMS
 CC CONTAIN 3 REPEATS WHILE TYPE II ISOFORMS CONTAIN 4 REPEATS.
 CC -!- PTM: PHOSPHORYLATION AT VARIOUS SERINE AND THREONINE RESIDUES IN
 CC S-P OR T-P MOTIFS BY PROLINE-DIRECTED PROTEIN KINASES (PDPK, CDC2,
 CC CDK5, GSK3, MAPK) (A FEW SITES PER PROTEIN IN INTERPHASE, MORE IN
 CC MITOSIS), AND AT SERINE RESIDUES IN K-X-G-S MOTIFS BY
 CC MAP/MICROTUBULE AFFINITY-REGULATING KINASE (MARK) (BY SIMILARITY).
 CC -!- DISEASE: MAY BE INVOLVED IN THE PATHOGENESIS OF CYTOPLASMIC
 CC INCLUSIONS (AS MALLORY BODIES) IN LIVERS OF MICE CHRONICALLY
 CC INTOXICATED WITH GRISOFULVIN OR DDC (3,5-DIETHOXYCARBONYL-2,4-
 CC DIHYDROCOLIDINE), A MODEL FOR HUMAN ALCOHOLIC HEPATITIS.
 CC ALTERATION OF TAU (ABNORMAL PHOSPHORYLATION AND CROSSLINKING)
 CC COULD CONTRIBUTE TO MALLORY BODIES FORMATION AND CROSSLINKING
 CC MICROTUBULE FUNCTION IN ALCOHOLIC LIVER DISEASE.
 CC -!- SIMILARITY: CONTAINS 4 TAU/MAP REPEATS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL; U12914; AAA58343.1; -
 DR EMBL; U12915; AAA58344.1; -
 DR EMBL; U12916; AAA58345.1; -
 DR EMBL; U12917; AAA58346.1; -
 DR EMBL; U12918; AAA58347.1; -
 DR EMBL; U12919; AAA58348.1; -
 DR EMBL; U12920; AAA58349.1; -
 DR EMBL; U12921; AAA58350.1; -
 DR EMBL; U12922; AAA58351.1; -
 DR EMBL; U12923; AAA58352.1; -
 DR EMBL; U12924; AAA58353.1; -
 DR EMBL; U12925; AAA58354.1; -
 DR EMBL; U12926; AAA58355.1; -
 DR EMBL; U12927; AAA58356.1; -
 DR EMBL; U12928; AAA58357.1; -
 DR EMBL; U12929; AAA58358.1; -
 DR EMBL; U12930; AAA58359.1; -
 DR EMBL; U12931; AAA58360.1; -
 DR EMBL; U12932; AAA58361.1; -
 DR EMBL; U12933; AAA58362.1; -
 DR EMBL; U12934; AAA58363.1; -
 DR EMBL; U12935; AAA58364.1; -
 DR EMBL; U12936; AAA58365.1; -
 DR EMBL; U12937; AAA58366.1; -
 DR EMBL; U12938; AAA58367.1; -
 DR EMBL; U12939; AAA58368.1; -
 DR EMBL; U12940; AAA58369.1; -
 DR EMBL; U12941; AAA58370.1; -
 DR EMBL; U12942; AAA58371.1; -
 DR EMBL; U12943; AAA58372.1; -
 DR EMBL; U12944; AAA58373.1; -
 DR EMBL; U12945; AAA58374.1; -
 DR EMBL; U12946; AAA58375.1; -
 DR EMBL; U12947; AAA58376.1; -
 DR EMBL; U12948; AAA58377.1; -
 DR EMBL; U12949; AAA58378.1; -
 DR EMBL; U12950; AAA58379.1; -
 DR EMBL; U12951; AAA58380.1; -
 DR EMBL; U12952; AAA58381.1; -
 DR EMBL; U12953; AAA58382.1; -
 DR EMBL; U12954; AAA58383.1; -
 DR EMBL; U12955; AAA58384.1; -
 DR EMBL; U12956; AAA58385.1; -
 DR EMBL; U12957; AAA58386.1; -
 DR EMBL; U12958; AAA58387.1; -
 DR EMBL; U12959; AAA58388.1; -
 DR EMBL; U12960; AAA58389.1; -
 DR EMBL; U12961; AAA58390.1; -
 DR EMBL; U12962; AAA58391.1; -
 DR EMBL; U12963; AAA58392.1; -
 DR EMBL; U12964; AAA58393.1; -
 DR EMBL; U12965; AAA58394.1; -
 DR EMBL; U12966; AAA58395.1; -
 DR EMBL; U12967; AAA58396.1; -
 DR EMBL; U12968; AAA58397.1; -
 DR EMBL; U12969; AAA58398.1; -
 DR EMBL; U12970; AAA58399.1; -
 DR EMBL; U12971; AAA58400.1; -
 DR EMBL; U12972; AAA58401.1; -
 DR EMBL; U12973; AAA58402.1; -
 DR EMBL; U12974; AAA58403.1; -
 DR EMBL; U12975; AAA58404.1; -
 DR EMBL; U12976; AAA58405.1; -
 DR EMBL; U12977; AAA58406.1; -
 DR EMBL; U12978; AAA58407.1; -
 DR EMBL; U12979; AAA58408.1; -
 DR EMBL; U12980; AAA58409.1; -
 DR EMBL; U12981; AAA58410.1; -
 DR EMBL; U12982; AAA58411.1; -
 DR EMBL; U12983; AAA58412.1; -
 DR EMBL; U12984; AAA58413.1; -
 DR EMBL; U12985; AAA58414.1; -
 DR EMBL; U12986; AAA58415.1; -
 DR EMBL; U12987; AAA58416.1; -
 DR EMBL; U12988; AAA58417.1; -
 DR EMBL; U12989; AAA58418.1; -
 DR EMBL; U12990; AAA58419.1; -
 DR EMBL; U12991; AAA58420.1; -
 DR EMBL; U12992; AAA58421.1; -
 DR EMBL; U12993; AAA58422.1; -
 DR EMBL; U12994; AAA58423.1; -
 DR EMBL; U12995; AAA58424.1; -
 DR EMBL; U12996; AAA58425.1; -
 DR EMBL; U12997; AAA58426.1; -
 DR EMBL; U12998; AAA58427.1; -
 DR EMBL; U12999; AAA58428.1; -
 DR EMBL; U13000; AAA58429.1; -
 DR EMBL; U13001; AAA58430.1; -
 DR EMBL; U13002; AAA58431.1; -
 DR EMBL; U13003; AAA58432.1; -
 DR EMBL; U13004; AAA58433.1; -
 DR EMBL; U13005; AAA58434.1; -
 DR EMBL; U13006; AAA58435.1; -
 DR EMBL; U13007; AAA58436.1; -
 DR EMBL; U13008; AAA58437.1; -
 DR EMBL; U13009; AAA58438.1; -
 DR EMBL; U13010; AAA58439.1; -
 DR EMBL; U13011; AAA58440.1; -
 DR EMBL; U13012; AAA58441.1; -
 DR EMBL; U13013; AAA58442.1; -
 DR EMBL; U13014; AAA58443.1; -
 DR EMBL; U13015; AAA58444.1; -
 DR EMBL; U13016; AAA58445.1; -
 DR EMBL; U13017; AAA58446.1; -
 DR EMBL; U13018; AAA58447.1; -
 DR EMBL; U13019; AAA58448.1; -
 DR EMBL; U13020; AAA58449.1; -
 DR EMBL; U13021; AAA58450.1; -
 DR EMBL; U13022; AAA58451.1; -
 DR EMBL; U13023; AAA58452.1; -
 DR EMBL; U13024; AAA58453.1; -
 DR EMBL; U13025; AAA58454.1; -
 DR EMBL; U13026; AAA58455.1; -
 DR EMBL; U13027; AAA58456.1; -
 DR EMBL; U13028; AAA58457.1; -
 DR EMBL; U13029; AAA58458.1; -
 DR EMBL; U13030; AAA58459.1; -
 DR EMBL; U13031; AAA58460.1; -
 DR EMBL; U13032; AAA58461.1; -
 DR EMBL; U13033; AAA58462.1; -
 DR EMBL; U13034; AAA58463.1; -
 DR EMBL; U13035; AAA58464.1; -
 DR EMBL; U13036; AAA58465.1; -
 DR EMBL; U13037; AAA58466.1; -
 DR EMBL; U13038; AAA58467.1; -
 DR EMBL; U13039; AAA58468.1; -
 DR EMBL; U13040; AAA58469.1; -
 DR EMBL; U13041; AAA58470.1; -
 DR EMBL; U13042; AAA58471.1; -
 DR EMBL; U13043; AAA58472.1; -
 DR EMBL; U13044; AAA58473.1; -
 DR EMBL; U13045; AAA58474.1; -
 DR EMBL; U13046; AAA58475.1; -
 DR EMBL; U13047; AAA58476.1; -
 DR EMBL; U13048; AAA58477.1; -
 DR EMBL; U13049; AAA58478.1; -
 DR EMBL; U13050; AAA58479.1; -
 DR EMBL; U13051; AAA58480.1; -
 DR EMBL; U13052; AAA58481.1; -
 DR EMBL; U13053; AAA58482.1; -
 DR EMBL; U13054; AAA58483.1; -
 DR EMBL; U13055; AAA58484.1; -
 DR EMBL; U13056; AAA58485.1; -
 DR EMBL; U13057; AAA58486.1; -
 DR EMBL; U13058; AAA58487.1; -
 DR EMBL; U13059; AAA58488.1; -
 DR EMBL; U13060; AAA58489.1; -
 DR EMBL; U13061; AAA58490.1; -
 DR EMBL; U13062; AAA58491.1; -
 DR EMBL; U13063; AAA58492.1; -
 DR EMBL; U13064; AAA58493.1; -
 DR EMBL; U13065; AAA58494.1; -
 DR EMBL; U13066; AAA58495.1; -
 DR EMBL; U13067; AAA58496.1; -
 DR EMBL; U13068; AAA58497.1; -
 DR EMBL; U13069; AAA58498.1; -
 DR EMBL; U13070; AAA58499.1; -
 DR EMBL; U13071; AAA58500.1; -
 DR EMBL; U13072; AAA58501.1; -
 DR EMBL; U13073; AAA58502.1; -
 DR EMBL; U13074; AAA58503.1; -
 DR EMBL; U13075; AAA58504.1; -
 DR EMBL; U13076; AAA58505.1; -
 DR EMBL; U13077; AAA58506.1; -
 DR EMBL; U13078; AAA58507.1; -
 DR EMBL; U13079; AAA58508.1; -
 DR EMBL; U13080; AAA58509.1; -
 DR EMBL; U13081; AAA58510.1; -
 DR EMBL; U13082; AAA58511.1; -
 DR EMBL; U13083; AAA58512.1; -
 DR EMBL; U13084; AAA58513.1; -
 DR EMBL; U13085; AAA58514.1; -
 DR EMBL; U13086; AAA58515.1; -
 DR EMBL; U13087; AAA58516.1; -
 DR EMBL; U13088; AAA58517.1; -
 DR EMBL; U13089; AAA58518.1; -
 DR EMBL; U13090; AAA58519.1; -
 DR EMBL; U13091; AAA58520.1; -
 DR EMBL; U13092; AAA58521.1; -
 DR EMBL; U13093; AAA58522.1; -
 DR EMBL; U13094; AAA58523.1; -
 DR EMBL; U13095; AAA58524.1; -
 DR EMBL; U13096; AAA58525.1; -
 DR EMBL; U13097; AAA58526.1; -
 DR EMBL; U13098; AAA58527.1; -
 DR EMBL; U13099; AAA58528.1; -
 DR EMBL; U13100; AAA58529.1; -
 DR EMBL; U13101; AAA58530.1; -
 DR EMBL; U13102; AAA58531.1; -
 DR EMBL; U13103; AAA58532.1; -
 DR EMBL; U13104; AAA58533.1; -
 DR EMBL; U13105; AAA58534.1; -
 DR EMBL; U13106; AAA58535.1; -
 DR EMBL; U13107; AAA58536.1; -
 DR EMBL; U13108; AAA58537.1; -
 DR EMBL; U13109; AAA58538.1; -
 DR EMBL; U13110; AAA58539.1; -
 DR EMBL; U13111; AAA58540.1; -
 DR EMBL; U13112; AAA58541.1; -
 DR EMBL; U13113; AAA58542.1; -
 DR EMBL; U13114; AAA58543.1; -
 DR EMBL; U13115; AAA58544.1; -
 DR EMBL; U13116; AAA58545.1; -
 DR EMBL; U13117; AAA58546.1; -
 DR EMBL; U13118; AAA58547.1; -
 DR EMBL; U13119; AAA58548.1; -
 DR EMBL; U13120; AAA58549.1; -
 DR EMBL; U13121; AAA58550.1; -
 DR EMBL; U13122; AAA58551.1; -
 DR EMBL; U13123; AAA58552.1; -
 DR EMBL; U13124; AAA58553.1; -
 DR EMBL; U13125; AAA58554.1; -
 DR EMBL; U13126; AAA58555.1; -
 DR EMBL; U13127; AAA58556.1; -
 DR EMBL; U13128; AAA58557.1; -
 DR EMBL; U13129; AAA58558.1; -
 DR EMBL; U13130; AAA58559.1; -
 DR EMBL; U13131; AAA58560.1; -
 DR EMBL; U13132; AAA58561.1; -
 DR EMBL; U13133; AAA58562.1; -
 DR EMBL; U13134; AAA58563.1; -
 DR EMBL; U13135; AAA58564.1; -
 DR EMBL; U13136; AAA58565.1; -
 DR EMBL; U13137; AAA58566.1; -
 DR EMBL; U13138; AAA58567.1; -
 DR EMBL; U13139; AAA58568.1; -
 DR EMBL; U13140; AAA58569.1; -
 DR EMBL; U13141; AAA58570.1; -
 DR EMBL; U13142; AAA58571.1; -
 DR EMBL; U13143; AAA58572.1; -
 DR EMBL; U13144; AAA58573.1; -
 DR EMBL; U13145; AAA58574.1; -
 DR EMBL; U13146; AAA58575.1; -
 DR EMBL; U13147; AAA58576.1; -
 DR EMBL; U13148; AAA58577.1; -
 DR EMBL; U13149; AAA58578.1; -
 DR EMBL; U13150; AAA58579.1; -
 DR EMBL; U13151; AAA58580.1; -
 DR EMBL; U13152; AAA58581.1; -
 DR EMBL; U13153; AAA58582.1; -
 DR EMBL; U13154; AAA58583.1; -
 DR EMBL; U13155; AAA58584.1; -
 DR EMBL; U13156; AAA58585.1; -
 DR EMBL; U13157; AAA58586.1; -
 DR EMBL; U13158; AAA58587.1; -
 DR EMBL; U13159; AAA58588.1; -
 DR EMBL; U13160; AAA58589.1; -
 DR EMBL; U13161; AAA58590.1; -
 DR EMBL; U13162; AAA58591.1; -
 DR EMBL; U13163; AAA58592.1; -
 DR EMBL; U13164; AAA58593.1; -
 DR EMBL; U13165; AAA58594.1; -
 DR EMBL; U13166; AAA58595.1; -
 DR EMBL; U13167; AAA58596.1; -
 DR EMBL; U13168; AAA58597.1; -
 DR EMBL; U13169; AAA58598.1; -
 DR EMBL; U13170; AAA58599.1; -
 DR EMBL; U13171; AAA58600.1; -
 DR EMBL; U13172; AAA58601.1; -
 DR EMBL; U13173; AAA58602.1; -
 DR EMBL; U13174; AAA58603.1; -
 DR EMBL; U13175; AAA58604.1; -
 DR EMBL; U13176; AAA58605.1; -
 DR EMBL; U13177; AAA58606.1; -
 DR EMBL; U13178; AAA58607.1; -
 DR EMBL; U13179; AAA58608.1; -
 DR EMBL; U13180; AAA58609.1; -
 DR EMBL; U13181; AAA58610.1; -
 DR EMBL; U13182; AAA58611.1; -
 DR EMBL; U13183; AAA58612.1; -
 DR EMBL; U13184; AAA58613.1; -
 DR EMBL; U13185; AAA58614.1; -
 DR EMBL; U13186; AAA58615.1; -
 DR EMBL; U13187; AAA58616.1; -
 DR EMBL; U13188; AAA58617.1; -
 DR EMBL; U13189; AAA58618.1; -
 DR EMBL; U13190; AAA58619.1; -
 DR EMBL; U13191; AAA58620.1; -
 DR EMBL; U13192; AAA58621.1; -
 DR EMBL; U13193; AAA58622.1; -
 DR EMBL; U13194; AAA58623.1; -
 DR EMBL; U13195; AAA58624.1; -
 DR EMBL; U13196; AAA58625.1; -
 DR EMBL; U13197; AAA58626.1; -
 DR EMBL; U13198; AAA58627.1; -
 DR EMBL; U13199; AAA58628.1; -
 DR EMBL; U13200; AAA58629.1; -
 DR EMBL; U13201; AAA58630.1; -
 DR EMBL; U13202; AAA58631.1; -
 DR EMBL; U13203; AAA58632.1; -
 DR EMBL; U13204; AAA58633.1; -
 DR EMBL; U13205; AAA58634.1; -
 DR EMBL; U13206; AAA58635.1; -
 DR EMBL; U13207; AAA58636.1; -
 DR EMBL; U13208; AAA58637.1; -
 DR EMBL; U13209; AAA58638.1; -
 DR EMBL; U13210; AAA58639.1; -
 DR EMBL; U13211; AAA58640.1; -
 DR EMBL; U13212; AAA58641.1; -
 DR EMBL; U13213; AAA58642.1; -
 DR EMBL; U13214; AAA58643.1; -
 DR EMBL; U13215; AAA58644.1; -
 DR EMBL; U13216; AAA58645.1; -
 DR EMBL; U13217; AAA58646.1; -
 DR EMBL; U13218; AAA58647.1; -
 DR EMBL; U13219; AAA58648.1; -
 DR EMBL; U13220; AAA58649.1; -
 DR EMBL; U13221; AAA58650.1; -
 DR EMBL; U13222; AAA58651.1; -
 DR EMBL; U13223; AAA58652.1; -
 DR EMBL; U13224; AAA58653.1; -
 DR EMBL; U13225; AAA58654.1; -
 DR EMBL; U13226; AAA58655.1; -
 DR EMBL; U13227; AAA58656.1; -
 DR EMBL; U13228; AAA58657.1; -
 DR EMBL; U13229; AAA58658.1; -
 DR EMBL; U13230; AAA58659.1; -
 DR EMBL; U13231; AAA58660.1; -
 DR EMBL; U13232; AAA58661.1; -
 DR EMBL; U13233; AAA58662.1; -
 DR EMBL; U13234; AAA58663.1; -
 DR EMBL; U13235; AAA58664.1; -
 DR EMBL; U13236; AAA58665.1; -
 DR EMBL; U13237; AAA58666.1; -
 DR EMBL; U13238; AAA58667.1; -
 DR EMBL; U13239; AAA58668.1; -
 DR EMBL; U13240; AAA58669.1; -
 DR EMBL; U13241; AAA58670.1; -
 DR EMBL; U13242; AAA58671.1; -
 DR EMBL; U13243; AAA58672.1; -
 DR EMBL; U13244; AAA58673.1; -
 DR EMBL; U13245; AAA58674.1; -
 DR EMBL; U13246; AAA58675.1; -
 DR EMBL; U13247; AAA58676.1; -
 DR EMBL; U13248; AAA58677.1; -
 DR EMBL; U13249; AAA58678.1; -
 DR EMBL; U13250; AAA58679.1; -
 DR EMBL; U13251; AAA58680.1; -
 DR EMBL; U13252; AAA58681.1; -
 DR EMBL; U13253; AAA58682.1; -
 DR EMBL; U13254; AAA58683.1; -
 DR EMBL; U13255; AAA58684.1; -
 DR EMBL; U13256; AAA58685.1; -
 DR EMBL; U13257; AAA58686.1; -
 DR EMBL; U13258; AAA58687.1; -
 DR EMBL; U13259; AAA58688.1; -
 DR EMBL; U13260; AAA58689.1; -
 DR EMBL; U13261; AAA58690.1; -
 DR EMBL; U13262; AAA58691.1; -
 DR EMBL; U13263; AAA58692.1; -
 DR EMBL; U13264; AAA58693.1; -
 DR EMBL; U13265; AAA58694.1; -
 DR EMBL; U13266; AAA58695.1; -
 DR EMBL; U13267; AAA58696.1; -
 DR EMBL; U13268; AAA58697.1; -
 DR EMBL; U13269; AAA58698.1; -
 DR EMBL; U13270; AAA58699.1; -
 DR EMBL; U13271; AAA58700.1; -
 DR EMBL; U13272; AAA58701.1; -
 DR EMBL; U13273; AAA58702.1; -
 DR EMBL; U13274; AAA58703.1; -
 DR EMBL; U13275; AAA58704.1; -
 DR EMBL; U13276; AAA58705.1; -
 DR EMBL; U13277; AAA58706.1; -
 DR EMBL; U13278; AAA58707.1; -
 DR EMBL; U13279; AAA58708.1; -
 DR EMBL; U13280; AAA58709.1; -
 DR EMBL; U13281; AAA58710.1; -
 DR EMBL; U13282; AAA58711.1; -
 DR EMBL; U13283; AAA58712.1; -
 DR EMBL; U13284; AAA58713.1; -
 DR EMBL; U13285; AAA58714.1; -
 DR EMBL; U13286; AAA58715.1; -
 DR EMBL; U13287; AAA58716.1; -
 DR EMBL; U13288; AAA58717.1; -
 DR EMBL; U13289; AAA58718.1; -
 DR EMBL; U13290; AAA58719.1; -
 DR EMBL; U13291; AAA58720.1; -
 DR EMBL; U13292; AAA58721.1; -
 DR EMBL; U13293; AAA58722.1; -
 DR EMBL; U13294; AAA58723.1; -
 DR EMBL; U13295; AAA58724.1; -
 DR EMBL; U13296; AAA58725.1; -
 DR EMBL; U13297; AAA58726.1; -
 DR EMBL; U13298; AAA58727.1; -
 DR EMBL; U13299; AAA58728.1; -
 DR EMBL; U13300; AAA58729.1; -
 DR EMBL; U13301; AAA58

```
465 sThrThrProSerProLysThrProProGlySerGlyCluProProLys 482
88 .....CGATCGGGTTTCAGTCCGCCGAGATCCGCTCGAAATAC 47
482 erGlyGluArgSerGlyTyrSerSerProGlySer..... 493
46 CTGCCTCTCGCCGAGGTACGGGATCGGTCTTCATCTCTGCCA 2
494 ...ProGlyThrProGlySerArgSerArgThrProSerLeuPro 507
```

seq_name: SwissProt_40:CC19_CAEEL

seq_documentation_block:

```
ID CC19_CAEEL STANDARD; PRT; 283 AA.
AC P18835;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Cuticle collagen 19.
GN COL-19.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=89326131; PubMed=2753356;
RA Cox G.N., Fields C., Kramer J.M., Rosenzweig B., Hirsh D.;
RT "Sequence comparisons of developmentally regulated collagen genes of
Caenorhabditis elegans."
RL Gene 76:331-344 (1989).
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS.
CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
CC COLLAGENS. COL-7, COL-8 AND COL-19 BELONGS TO THE SAME GROUP OF
CC COLLAGEN.
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
DR EMBL; M25481; AAA27987.1; -
DR PIR; JS0170; JS0170.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.
FT DOMAIN 95 124 TRIPLE-HELICAL REGION.
FT DOMAIN 141 263 TRIPLE-HELICAL REGION.
SQ SEQUENCE 283 AA; 27565 MW; FD53BE2785D1767C CRC64;
```

alignment_scores:

Quality:	90.50	Length:	82
Ratio:	1.847	Gaps:	6
Percent Similarity:	59.756	Percent Identity:	36.585

alignment_block:

US-09-462-480-4/rev x CC19_CAEEL ..

Align seg 1/1 to: CC19_CAEEL from: 1 to: 283

```
277 CCTGCTGCTCTCTCGCGCCTCGAGTATTGGACGCCGCGCTGACGA 228
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
178 ProGlnGlyAspAlaGlyArgProGlyAlaAlaGly...ArgProGlyPr 193
227 ATATTGCTGAGATCTCGTCGAGTCTCTGCTTCTGCTTATTGGCTGCTTC 178
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
193 oAlaGlyProArgGlyGlu.....ProGlyThrGluTyr..... 204
177 TTGGAAGCGCACACCGCGGCTGGCGGCGCTGCCCG.....CCGCGC 134
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
205 .....LysProGlyGlnProGlyArgProGlyProGlyProGlyProArg 218
133 CGCGCCACTGCGCTGCAACGAACCTGCGCTGCGACTCCACTGGTCGATC 84
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
219 GlyGluThrGlyProAlaGlyAsn...ProGlyAla...ProGlyAsnAs 233
83 TGGGTTTTTCAGTCTCGCGGAGATCGCTCGAAATACCTGCTCCT 38
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
233 pGlyGluAlaGlyLysAsnGlyAsnAlaGlyArgProGlyProPro 248
```

seq_name: SwissProt_40:WAS3_HUMAN

seq_documentation_block:

```
ID WAS3_HUMAN STANDARD; PRT; 502 AA.
AC Q9UPY6; O94974;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Wiskott-Aldrich syndrome protein family member 3 (WASP-family
DE protein member 3) (Verprolin homology domain-containing protein 3).
GN WASF3 OR WAVE3 OR SCAR3 OR KIAA0900.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99310608; PubMed=10381382;
RA Suetsugu S., Miki H., Takenawa T.;
RT "Identification of two human WAVE/SCAR homologues as general actin
RT regulatory molecules which associate with the Arp2/3 complex."
RL Biochem. Biophys. Res. Commun. 260:296-302(1999).
RN [2]
RP SEQUENCE OF 48-502 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 5:355-364(1998).
RN [3]
RP SEQUENCE OF 48-502 FROM N.A.
RA Machesky L.M., Insall R.H.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DOWNSTREAM EFFECTOR MOLECULES INVOLVED IN THE
CC TRANSMISSION OF SIGNALS FROM TYROSINE KINASE RECEPTORS AND SMALL
CC GTPASES TO THE ACTIN CYTOSKELETON.
CC -!- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN OVARY AND BRAIN.
CC -!- DOMAIN: BINDS ARP2/3 COMPLEX THROUGH THE C-TERMINAL REGION AND
CC ACTIN THROUGH VERPROLIN HOMOLGY (VPH) DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 WH2 DOMAIN.
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

OM of: US-09-462-480-4 to: SPTRMBL19:* out_format : pfs

Date: Jul 22, 2002. 1:40 AM

About: Results were produced by the GenCore software; version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+n2p.model -DEV=xlh  
-O=/cgn2_1/USPTO_spool/US09462480/runat.18072002.164419.19544/app_query.fasta_1.2850  
-DB=SPTRMBL19 -OPMT=fastan -SUFFIX=rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bsum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SOURCE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTENT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480 -CGN1_1.133 -NCPU=6 -ICPU3 -LONLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-462-480-4

Query length: 302

Database: SPTRMBL19:*

Database sequences: 562222

Database length: 172994929

Search time (sec): 143.350000

score_list:

Sequence	Strd	Orig	zscore	EScore	Len	Documentation
sp_plant:Q92W08	-	119.00	189.83	0.0035	891	Q92W08 arabidopsis thaliana (md
sp_plant:Q9FYB2	-	119.00	189.80	0.0035	894	Q9FYB2 arabidopsis thaliana (md
sp_human:Q94932	-	109.00	175.33	0.0314	645	Q94932 homo sapiens (human). ki
sp_rodent:Q61359	-	105.50	177.94	0.0597	243	Q61359 mus musculus (mouse). dy
sp_virus:Q9DWC9	-	104.00	175.01	0.0838	252	Q9DWC9 rat cytomegalovirus (str
sp_rodent:Q9CUM3	-	103.50	174.04	0.0939	255	Q9CUM3 mus musculus (mouse). 29
sp_bacteria:Q93K69	+	102.00	163.52	0.1488	620	Q93K69 microbacterium laevanifo
sp_rodent:Q93K47	+	102.00	157.65	0.1634	1198	Q93K47 rattus norvegicus (rat)
sp_rodent:Q70495	-	101.00	158.49	0.1960	897	Q70495 mus musculus (mouse). pl
sp_bacteria:Q87852	+	99.50	156.17	0.2721	859	Q87852 streptomyces coelicolor. r
sp_bacteriap:Q9RX36	+	99.50	156.13	0.2727	873	Q9RX36 deinococcus radiodurans. r
sp_vertebrate:Q57580	-	99.00	152.80	0.3171	1151	Q57580 gallus gallus (chicken). r
sp_plant:Q9XJ18	-	98.50	163.41	0.2951	317	Q9XJ18 oryza sativa (rice). sim
sp_plant:Q94EV7	-	98.50	161.20	0.3057	406	Q94EV7 zea mays (maize). map ki
sp_human:Q9U039	+	98.50	153.58	0.3452	956	Q9U039 homo sapiens (human). r
sp_human:Q9U040	+	98.50	151.11	0.3591	1262	Q9U040 homo sapiens (human). r
sp_human:Q9UHA8	+	98.50	145.78	0.3910	2296	Q9UHA8 homo sapiens (human). s
sp_human:Q9UQ35	+	98.50	144.16	0.4012	2752	Q9UQ35 homo sapiens (human). r
sp_invertebrate:Q952N1	+	97.50	167.43	0.3364	166	Q952N1 caenorhabditis elegans
sp_invertebrate:Q952N2	+	97.50	165.42	0.3474	208	Q952N2 caenorhabditis elegans
sp_invertebrate:Q9WOC8	-	97.50	162.19	0.3658	299	Q9WOC8 drosophila melanogast
sp_rodent:Q61360	-	97.50	155.81	0.4049	612	Q61360 mus musculus (mouse). dy
sp_invertebrate:Q95T35	-	97.50	149.23	0.4498	1282	Q95T35 drosophila melanogast
sp_bacteria:Q9FC14	-	96.50	159.00	0.4243	388	Q9FC14 streptomyces coelicolor. r
sp_plant:Q9FER8	-	96.50	164.44	0.4289	191	Q9FER8 zea mays (maize). hmg1/y
sp_bacteria:Q9S2Y7	-	96.50	155.72	0.4930	509	Q9S2Y7 streptomyces coelicolor. r
sp_human:Q9P2P0	-	96.00	147.56	0.6190	1154	Q9P2P0 homo sapiens (human). k
sp_bacteria:Q9LBV2	-	95.50	159.40	0.5651	277	Q9LBV2 streptomyces virginiae. r
sp_bacteria:Q30894	+	95.50	158.48	0.5734	307	Q30894 myxococcus xanthus. hypc
sp_archaeap:Q9YD41	-	95.00	166.20	0.5589	117	Q9YD41 aeropyrum pernix. hypot
sp_invertebrate:Q966L6	-	95.00	153.20	0.6877	504	Q966L6 caenorhabditis elegans
sp_plant:Q9FY55	-	94.50	160.88	0.6708	193	Q9FY55 zea mays (maize). high m
sp_invertebrate:Q9NFZ9	-	94.50	157.56	0.7072	280	Q9NFZ9 ascaris suum (pig round
sp_invertebrate:Q17626	+	94.00	159.82	0.7521	197	Q17626 caenorhabditis elegans
sp_plant:Q9SE9E	-	94.00	153.21	0.8358	414	Q9SE9E arabidopsis thaliana (md
sp_plant:Q94AS1	-	94.00	153.21	0.8358	414	Q94AS1 arabidopsis thaliana (md
sp_virus:Q72759	-	93.50	138.62	1.16	1933	Q72759 cowpox virus (cpv). b22
sp_vertebrate:Q91810	-	93.00	152.26	1.03	379	Q91810 xenopus laevis (afrikan
sp_plant:Q06804	-	92.50	160.99	0.9894	129	Q06804 hordeum vulgare (barley)
sp_human:Q96030	-	92.50	156.23	1.07	220	Q96030 homo sapiens (human). ad

seq_name: sp_plant:Q92W08

seq_documentation_block:

ID Q92W08 PRELIMINARY; PRT; 891 AA.

AC Q92W08;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE PUTATIVE PROLINE-RICH PROTEIN.

GN AT2G29210.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RX MEDLINE=20083487; PubMed=10617197;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,

RA Cronin L.A., Shen M., VanAken S.E., Unayam L., Tallon L.J., Gill J.E.,

RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,

RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;

RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis

thaliana";

RL Nature 402:761-768(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC004561; AAC95214.1; -

DR InterPro; IPR002483; PWI.

DR InterPro; IPR002965; P-rich_extens.

DR Pfam; PF01480; PWI; 1.

DR PRINTS; PR01217; PRICHEXTENS.

DR SMART; SM00311; PWI; 1

SQ SEQUENCE 891 AA; 102261 MW; C0CCB6DE05E528B CRC64;

alignment_scores:

Quality: 119.00 Length: 111

Ratio: 2.052 Gaps: 6

Percent Similarity: 52.252 Percent Identity: 35.135

alignment_block:

US-09-462-480-4/rev x Q92W08 ..

Align seg 1/1 to: Q92W08 from: 1 to: 891

301 AGAGCCCATTTTGGAGACAGCGCTGCTGCTCTCT...CGTCGGCC 255

|||||:|||||:|||||: ||: ||: ||: ||| ||| ||

425 ArgSerProLeuAlaLysArgGlyArgSerAspSerProGlyArgSerPr 441

254 CTCGAGTATTGGAGCCCGCTCGAGATATTCGTCGAGA..... 215

|||||:|||||:|||||: |||: |||: |||: |||

441 oSerProValAlaArgLeuArgAspProThrGlyAlaArgLeuProSerP 458

214TCFCGTCGAGTCTCTGCTTCTGCTTATTGCTTCTTCTTGA 173

|||||:|||||:|||||: |||: |||: |||: |||

458 roSerIleGluGlnArgLeuProSerProProValAlaGlnArgLeuPro 474

172 AGCGACCAACCG...CGCCCTGGGCGCGCTCCCGCGCGCGCGCCAC 126

||| ||||| ||| ||| ||||| |||||

```

75 CAGGTGCGCCGGAGA.....TCCGCTGAAATTACCTG..... 44
   ::      |||      |||      |||      |||
506 roSerProMetArgfilecglySerHisAlaAlaAsnHisLeuGluSer 522

43 .....CCTCTGCCCCGA 32
   |||||::|||
523 ProSerProSerLeuSerProProGlyArg 533

seq_name: sp_human:O94932

seq_documentation_block:
ID O94932 PRELIMINARY; PRT; 645 AA.
AC O94932;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA0847 PROTEIN (FRAGMENT).
GN KIAA0847.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hiroseawa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT xii. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro. ";
RL DNA Res. 5:355-364(1998).
DR EMBL; AB020654; BAA74870.1; -.
FT NON_TER 1
SQ SEQUENCE 645 AA; 70370 MW; 282EE9ABAB37ADFE CRC64;

```

```

alignment_scores:
  Quality: 109.00      Length: 106
  Ratio: 1.817        Gaps: 3
  Percent Similarity: 56.604  Percent Identity: 30.189

alignment_block:
  US-09-462-480-4/rev x 094932  ..

  Align seg 1/1 to: 094932  from: 1  to: 645

301 AGAAGCCCATTTGGCGAGGACGGCTGCTGCTCTCTCGCGCCCTC 252
   :::::::::::  ::::  ::::  ::::  ::::  ::::  ::::
530 LysSerProGlySerProHisAsnProLysThrProProLysSerPro.. 545

251 GAGTATTGACGCGCGCCCTGACGAATATTGCTCGAGATCTGTCGATTC 202
   ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
546 _ValValProArgArgSerProSerAlaSerProArgSerSerLeup 562

201 CTGCTTCTGCTATTATGGTGCTTCTTTGGAAGCGCACACCGCGCGCTGG 152
   ||  :::::  ::::  ::::  ::::  ::::  ::::  ::::
562 roArgThrSer .....SerSerProSerArgAlaGly 573

151 CGGCGCTCCCGCGCGCGCGCATCTGGCCCTGCAACGACCTGCCCTC 102
   ||||||  ::::  ::::  ::::  ::::  ::::  ::::
574 ArgProHisHisAspGlnArgSerSerSerProHisLeuGlyArgSerLys 590

101 GACTCCACTGTGTCATCTGGGTTTTCAGTCGCGCGGAGATCC..... 59
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
590 sSerProProSerHisSerGlySerSerSerSerArgArgSerCysGlnG 607

58 .....GCTCGAAATTACTGCTCTCTGCCCGGAGGTAG 26
   ::::  ::::  ::::  ::::  ::::  ::::
607 lnGluHisCysLysProSerLysAsnGlyLeuLysGlySerGlySerLeu 623

25 CGGCGTGGCTCTTCATCT 8
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

[illegible]

GN 2900057H02RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK013719; BAB28968.1; ..
DR MGD; MGI:1924225; 2900057H02RIK.
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 1
SQ SEQUENCE 255 AA; 27606 MW; 07F5DE6B902D5FE6 CRC64;

alignment_scores:
Quality: 103.50 Length: 116
Ratio: 1.754 Gaps: 5
Percent Similarity: 50.862 Percent Identity: 31.897

alignment_block:

US-09-462-480-4/rev x Q9CUW3

Align seg 1/1 to: Q9CUW3 from: 1 to: 255

283 ACACGGCT.....GCTGCTGCTCTGCTGGCGCCCTGAGTATTGGAGC 240
|||||
40 ThrAlaProGlyGlnAlaAlaGlyProGlyArgProGlnGlnProArgG1 56
239 C.....CGGCTGACGAATATTCGTGCA 217
nGlyAspProGlyProAlaGlyProHisArgProArgAlaPro**GlnA 73
216 GATCTGCTCGAGTTCCTGCTCTGCTATTGCTGCTTCTTGGAGCGCA 167
|||||
73 rGluGlyArgAlaProGlyPro.....GluAla 82
166 CCACCGGGCTGGCGGGCGCTGCCG....CCGCGCGCGCCACTGGCC 120
|||||
83 ProAlaGlyProGlyGlnProProProProProProProProProProGly 99
119 TGCAACGAACCTGCGCTGACT..... 98
99 uSerThrArgProThrProThrProGlyArgGlyProGlyGluThrPro 116
97CCACCTGGTGATCTGGGTTTCAGTTCGCGCGAGA 62
116 erGly**GlyArgProProGlyArgLysGluGlyThrArgAlaArg 132
61 TCCGCTGGAATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14
::: |||||

133 AspProArgLysProArgArgAlaLysSerSerArgArgSer 148
seq_name: sp_bacteria:Q93R69

seq_documentation_block:

ID Q93R69 PRELIMINARY; PRT; 620 AA.

AC Q93R69;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LEVANASE.
OS Microbacterium laevaniformans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococciaceae; Microbacteriaceae; Microbacterium.
OX NCBI_TaxID=36807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15953;
RA Song E., Cha J.;
RT "Microbacterium laevaniformans levanase-complete cds."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB062062; BAB59060.1; ..
SQ SEQUENCE 620 AA; 67667 MW; 82B4909328CBAAC2 CRC64;

alignment_scores:

Quality: 102.00 Length: 105
Ratio: 1.821 Gaps: 5
Percent Similarity: 53.333 Percent Identity: 34.286

alignment_block:

US-09-462-480-4 x Q93R69

Align seg 1/1 to: Q93R69 from: 1 to: 620

29 CCCTGGCGCAGGAGG.....CAGGTATTTCGACGGGATCTCCGGCGAC 72
|||||
520 ProSerGlyAspGlnGlyLeuAlaLeuThrSerGlyGlyProAlaVa 536
73 CTGAAA.....CCCA 83
536 lPheArgAspLeuArgIleThrGluPheAlaAspLeuAlaHisValProA 553
84 GA...TCGACCAAGTTCGAGTCGACGGCAGGTCGTTGTCAGGGCCAGTGCC 130
|| |||||
553 rGAlaSerSerArgThrSer.....ArgAlaThrArg 563
131 CGCGCGCGCGGCGACCGCGCCCGCGCGGTGGTGGTTCCTCCAGAA 180
:::||||| |||||
564 GlyAlaArgGlyGlyArgProProAlaProSerProGlySerValArgAr 580
181 GCACCCAATAGCAGACGAGGAAGTCTG.....ACGAGATCTC 218
::: ||| ::|||
580 gArgAlaIleCysAlaGlyArgArgAlaArgAlaTrpProThrArgSerP 597
219 GACCAATATTCGTCAGCGCGGCTCCCAATCTCGAGCGCGCGCAGCAGGAGC 268
|||||
597 rGlyArgThrProArgProAlaSerSerArgArgArgProSerArgSer 613
269 AGCAGCAGCGCTGT 283
614 ThrGlyThrArgCys 618

seq_name: sp_rodent:Q9JKA7

seq_documentation_block:

ID Q9JKA7 PRELIMINARY; PRT; 1198 AA.

AC Q9JKA7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPERPOLARIZATION-ACTIVATED, CYCLIC NUCLEOTIDE-GATED POTASSIUM CHANNEL
DE 4.

SEQUENCE FROM N.A.
STRAIN=A3(2);
Seeger K.J., Harris D.;
submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
STRAIN=A3(2);
Parkhill J., Barrell B.G., Rajandream M.A.;
submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
STRAIN=A3(2);
MEDLINE=97000351; PubMed=8843436;
Redenbach M., Kleser H.M., Denapite D., Eichner A., Cullis
K., Kashi H., Hopwood D.A.;
"A set of ordered cosmids and a detailed genetic and physical
map of the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
Mol. Microbiol. 21:77-96(1996).
EMBL; AL031013; CAA19790.1;
InterPro; IPR001764; Glyco_hydro_3.
InterPro; IPR002772; Glyco_hydro_3C.
Pfam; PF00933; Glyco_hydro_3; 1.
Pfam; PF01915; Glyco_hydro_3; 1.
PRINTS; PR00133; GLHYDRLASE3.
SEQUENCE 859 AA; 92455 MG;
SQ

alignment_scores:		
Quality:	99.50	Length:
Ratio:	2.117	Gaps:
Percent Similarity:	51.087	Percent Identity:

alignment_block:

US-09-462-480-4 x 087852

Align seg 1/1 to: 087852 from: 1 to: 859

```

39  GGAGGCGAGTAAATTCGAGCGGATCTCCGCGACCTGAAAACCCACGATCG 88
287 GlyGlyArg.....aspValSerArgSerProvalProGlyArg 199
89  ACCAGGT.....GGATCGACGCGAGGTTCGTTGCAG 120
299  gproAlaAspSerProGlyAlaGluGlyAlaaspGlyGlyAlaGlyAlaG 316
121  GCGCAGTGGCGGCGCGGGGACGGCGGCCCGACGGCGCGGT.....164
316  lProSerSer.....GlyAlaGluGlyLeuProGlyArgGlyProAla 330
165  ...GGTCGCGCTTCCAAGAGCAGGCCAAATAAGCAGAGCAGGAACTCGAGC 211
331  HisGlyAlaLysProSerGlyProArgProArgAlaGlyAspGlyArg 347
212  AGATCTCGACGAATATTCGTGAGCGCGCGCTCCAATACTCGAGGSCCGAC 261
347  gAlaLeu.....AlaArgA 352
262  GAGGAGCAGCAGCAGGCGCTGCTCTC 287
352  rGAlaValAlaAlaGlyAlaValLeu 360

```

seq_name: sp_bacteriap:O9RX36

seq_documentation_block:

ID	Q9RX36	PRELIMINARY;	PRT;	873 AA.
AC	Q9PV36.			

AC Q9RX36;
DT 01-MAY-2000 (TREMBlere) 13 Created

DT	01-MAY-2000	(TREMBLrel. 13, Last seq
DT	01-MAY-2000	(TREMBLrel. 13, Last seq

DT 01-JUN-2001 (TRENBLER, 17, Last annotation update)
DE PENICILLIN-BINDING PROTEIN 1

GN DR0479. PENICILLIN-BINDING PROTEIN 1.

OS *Deinococcus radiodurans*.

Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.

NCBI_TaxID=1299;	
[1]	
SEQUENCE FROM N.A.	
STRAIN=R1;	
MEDLINE=20036896; PubMed=10567266;	
White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,	
Moffat R.J., Hart D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,	
Doizat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,	
Vamatheva J.J., Lam P., McDonald L., Utterback T., Zalewski C.,	
Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,	
Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,	
Fraser C.M.;	
"Genome sequence of the radioresistant bacterium Deinococcus	
radiodurans R1.";	
Science 286:1571-1577(1999).	
EMBL; AE001907; AAF10059.1; -.	
TIGR; DR0479; -.	
InterPro: IPR002965; P_rich_extensn.	
InterPro: IPR001264; Transglycosyl.	
InterPro: IPR001460; Transpeptidase.	
Pfam; PF00912; Transglycosyl; 1.	
Pfam; PF00905; Transpeptidase; 1.	
PRINTS; PR01217; PRICHEXTENS.	
ProDom; PD001895; Transglycosyl; 1.	
Complete proteome.	
SEQUENCE 873 AA; 95571 MW; 8B859AC6B9919E086 CRC64;	

alignment_scores:

Quality: 99.50

Percent Similarity:	42.222	Percent Similarity:	42.222
Ratio:	1.746	Ratio:	1.746
Gaps:	2	Gaps:	2
Length:	133	Length:	133

Percent Similar

```
alignment_block:
```

US-09-462-480-4/rev x Q9RX36 ..

```

Align seg 1/1 to: Q9RX36 from: 1 to: 873

298 AGCCCATTTGCGAGGACAGCGCTGCTGCTCTGCTGCGCCCTCGAG 249
    :||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
724 ThrProAlaValThrAlaPro.....GluProGlnProAlaIath 738

248 T.....ATTGGACGCGCGCTTCAGCAATATTTCGTGGAGATCTGCTC 208
    : ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
738 rThrThrProAlaGlyArgSerProAlaProLeuArgArgThrArgA 755

207 GAGTT.....CCTGCTTCTGCTATTGGCTGCTCTT 176
    |||: ||| ||| ||| ||| ||| ||| ||| ||| |||
755 rgThrLeuProArgGlnThrSerProArgArgAla.....Leu 767

175 GGAAGCGCACACCGCGGCTGGGGCGG...TCCCGCGCG..... 137
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
768 ProSerArgThrCysArgProArgArgProLeuSerProProArgAr 784

136 .....CGCGCGCCACTGCCCT. 119

784 gProArgProGlyAlaThrArgArgAlaAsnArgArgThrSerProp 801
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 .....GCAACGAACCTGCCCTCG..... 101
    |||||: |||: |||: |||: |||: |||: |||: |||: |||:
801 roSerAlaThrArgCysProThrCysProArgArgArgLeuSerArgCys 817

100 .....ACTCACTGGTCG 87
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
818 ProAlaThrProArgProLeuSerArgArgProValThrProProGlyAr 834

86 ATCTGGGTTTACAGTCGCGGAGATCCGCTCGCAATTACTGCCCTCGT 37
    | |||||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
834 gPro...PheGlnAlaArgArgArgIleProArgLeuArgIleProProt 850
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

36 CCCGA 32

850 brara 957
::|||

850 hrArg 851

seq_name: sp_vertibrate:057580

```
seq_documentation_block:
ID   O57580      PRELIMINARY;          PRT;   1151 AA.
AC   O57580;
DT   01-JUN-1998 (TREMBLrel. 06, Created)
DT   01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE   HIGH MOLECULAR MASS NUCLEAR ANTIGEN (FRAGMENT).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98034404; PubMed=9365273;
RA   Shimada K., Harata M., Mizuno S.;
RT   "A nuclear matrix-associated high molecular mass nuclear antigen,
RT   HMNA, of chicken and marked decrease of its immunoreactivity during
RT   the progression of S phase."
RL   J. Cell Sci. 110:3031-3041(1997).
DR   EMBL; D88440; BAA24137.1; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   PRINTS; PR001778; POA_allergen.
DR   PRINTS; PR00015; GPOSANCHOR.
DR   PRINTS; PR00833; POAALLERGEN.
FT   NON_TER 1
FT   NON_TER 1151
FT   NON_TER 1151
SQ   SEQUENCE 1151 AA; 109708 MW; 2403F6835F9A2AB3 CRC64;
```

```
alignment_scores:
Quality: 99.00      Length: 115
Ratio: 2.062       Gaps: 4
Percent Similarity: 41.739   Percent Identity: 26.957

alignment_block:
US-09-462-480-4/rev x O57580 ..

Align seg 1/1 to: O57580 from: 1 to: 1151

298 AGCCATTGGAGGACAGCCGCTGCTCTGCTCGCGCCCTCGAG 249
   :::::
16 AlaProSerProAlaProAlaProAlaProAlaProAlaPro..... 30

248 TATTGGAGCGCGCTGACGAATATTCGTCGAGATCTCGTAGATTCTGT 199
   ||| ::| |||||
31 .....ProArgProLysTrpValProI 38

198 CTTCGCTTAT.....TGGCTGCTT 179
   :::::
38 leAlaGluLeuH1ProAlaAlaProGlnProProLysTrpValPro 54

178 CTTGGAAGCGCACCGCGCTGGG...CGGCGCTCCCGCGCGCGCG 132
   :::::||||||| ||||| :::::||||||| ::
55 ileGlyGlyAlaProProProGlyThrGluProThrProProSerL 71

131 C.....GCCACTG 124
   :
71 sProThrAspGlyAlaAspAlaAlaProLysAlaSerAlaGluLeuThrs 88

123 GCCCTGACCAACCTGCCCTGACTCCACCTGGTCGATCTGGGTTTCA 74
   ::||| ||||| ||| ::::: |||
88 erProProAlaSerProProAspGlyProLysAlaProSer 104

73 GGTGCGCGGAGATCGCTCGAAATACCTGCTCTCGCGCGAGG 29
   ||| ::::: ::| ||||| ::::: |||
105 GlyAlaGlyGluAlaGluAlaGlyThrProProSerGingly 119
```

seq_name: sp_plant:O9XJ18

```
seq_documentation_block:
ID   O9XJ18      PRELIMINARY;          PRT;   317 AA.
AC   O9XJ18;
DT   01-NOV-1999 (TREMBLrel. 12, Created)
DT   01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE   SIMILAR TO HUMAN MRNA FOR ALPHA 1.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzaceae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Nagamura Y., Yamamoto K.;
RT   "Oryza sativa nipponbare(CA3) genomic DNA, chromosome 6, PAC
RT   clone:P0680A03."
RL   Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR   EMBL; AB023482; BAA78734.1; -.
SQ   SEQUENCE 317 AA; 34259 MW; 060873B38658BB0A CRC64;

alignment_scores:
Quality: 98.50      Length: 144
Ratio: 1.698       Gaps: 8
Percent Similarity: 40.278   Percent Identity: 31.944

alignment_block:
US-09-462-480-4/rev x O9XJ18 ..

Align seg 1/1 to: O9XJ18 from: 1 to: 317

280 GCGCTGCTGCTCTCTCTCGCGCCCTCGAGTATTGACCGCGCCTGA 231
   ||||| ||::: ||||| ::::: |||
48 AlaPro...AlaGlyGluArgArgGlyAlaAlaAspGlyGlyTrpGluG 63

230 CGAATATTTCGAGATCTCGTCTCGAGTTCCTCTCTCTCTTAT..... 188
   :||| ||::: ||| ||||| ::|
63 uGluArgSerAlaArgTyrArgArgIleMetArgGlyArgAsnGlyHis 80

187 .....TGGCTGCTTCTGGAGCGCACCGCGCGCTGGGGG..... 149
   ||||| ||||| |||||
80 yrProTrpLeuLeu.....ProProThrGlyAlaAlaArgAlaPro 93

149 ..... 149

94 LeuLeuSerArgArgSerProLeuMetSerProLeuProLeuGlySerTh 110

148 ....CCGTCCCGCGCGCGCG.....GCCA 127
   ||| ||||| |||||
110 rThrProLeuProProArgArgGlyGlnArgLeuArgArgHisSerSerH 127

126 CTGCGCCCTGCAACGAACTCGCTCGACTCCACCT..... 92
   ||||| ||||| |||||
127 isProProAlaProLeuHisProSerThrProProGlyHisArgGlySer 143

91 .....GGTCGATCTGGGTTTTCAGTCGCGGAGATCCGCTCGAAA 51
   ||||| ||| ::| |||||
144 GlySerArgGlyArgArgGlyCysGlyArgLeuLeuAlaProLeuPr 160

50 TTACTGCTCTCT.....GCCCGAGGTAGCGGC 22
   ||||| ||| ||| |||
160 oLeuLeuProProThrLysThrThrThrMetAlaThrGlySerArgA 177

21 ATCGS.....TCTTCATCTCTGCCA 2
   ::||| ||||| |||
177 rgArgArgThrArgThrSerSerSerGlnPro 187

seq_name: sp_plant:O94EV7

seq_documentation_block:
ID   O94EV7      PRELIMINARY;          PRT;   406 AA.
```

```
AC Q94EV7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MAP KINASE KINASE.
GN MAPKK1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI|TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSOME-HUCK1;
RA Fu H., Zheng Z., Dooner H.K.;
RT "Large differences in recombination rates within adjacent gene-dense
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RW EMBL; AF391808; AAK73104.1;
KW Kinase.
SQ SEQUENCE 406 AA; 44664 MW; DBFIDE6568F47DF0 CRC64;

alignment_scores:
  Quality: 98.50      Length: 100
  Ratio: 1.894       Gaps: 5
  Percent Similarity: 52.000   Percent Identity: 38.000

alignment_block:
US-09-462-480-4/rev x Q94EV7
..
Align seg 1/1 to: Q94EV7 from: 1 to: 406

286 AGGACAGCGCTGCTGCT...GCTCCTCGTCGCGCCTCGAGTATTGGACG 240
|||||:|||||:||||| ||| |||||||||:|||||
56 ArgThrProCysAlaAraAlaAlaSerArgProSerProAlaAar 72
|||||:|||||:||||| ||| |||||||||:|||||
239 CGGCGCTGACGAATATTCGTGCGA.....T 214
|||||:|||||:||||| ||| |||||||||:|||||
72 gArgAlaAlaArgAlaAlaAlaArgIleSerProSerProCysProSerA 89
|||||:|||||:||||| ||| |||||||||:|||||
213 CTCGTCGAGTCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 164
|||||:|||||:||||| ||| |||||||||:|||||
89 IaArgSerProProProSerProSerArgSerLeuSerArgGlnProPro 105
|||||:|||||:||||| ||| |||||||||:|||||
163 dCGCGGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCAAC 114
|||||:|||||:||||| ||| |||||||||:|||||
106 ProProGly.....GlyProProArgAlaSer..... 116
|||||:|||||:||||| ||| |||||||||:|||||
113 GAACCTGCGCTGACTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 64
|||||:|||||:||||| ||| |||||||||:|||||
117 .....ProArgProProProArgSerSerSerSerSerSerSerSerProp 131
|||||:|||||:||||| ||| |||||||||:|||||
63 GATCGCTCGAATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 16
|||||:|||||:||||| ||| |||||||||:|||||
131 roLeuAla.GluLeuGluArgValArgValGlySerGlyAlaGly 146

seq_name: sp_human:Q9UQ39
seq_documentation_block:
ID Q9UQ39 PRELIMINARY; PRT; 956 AA.
AC Q9UQ39;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RNA BINDING PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI|TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.
```

```
RT "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016088; BAA83714.1;
DR InterPro; IPR002965; P_Rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS1.
FT NON_TER 1
FT NON_TER 956
SQ SEQUENCE 956 AA; 106462 MW; C9FEAD07B49F5011 CRC64;

alignment_scores:
  Quality: 98.50      Length: 140
  Ratio: 1.669       Gaps: 5
  Percent Similarity: 42.143   Percent Identity: 27.143

alignment_block:
US-09-462-480-4 x Q9UQ39
..
Align seg 1/1 to: Q9UQ39 from: 1 to: 956

14 AGACCGATGCGCTACCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 55
:||||| ||| |||:|||||:|||||:||||| |||
407 LysProAlaProAlaProGlySerHisArgGluIleSerSerProTh 423
55 .....
423 rSerLysAsnArgSerHisGlyArgAlaLysArgAspLysSerHisSerH 440
56 .....ACGGATCTCCGCGCGACCTGAAAA 79
|||||:|||||:||||| ||| |||||||||:|||||
440 IsThrProSerArgArgMetGlyArgSerArgSerProAlaThrAlaLys 456
80 .....CC 81
457 ArgGlyArgSerArgSerArgThrProThrLysArgGlyHisSerArgSe 473
82 CAGATCGACACAGTGGAGTCGA.....CGCAGGTTCGTTGCAGGG 122
|||||:|||||:||||| ||| |||||||||:|||||
473 rArgSerProGlnTrpArgArgSerArgSerAlaGlnArgTrpGlyArgS 490
123 CCAGTGGCGCGCGCGCGGAGCGG.....CGCCCGAGCGCGCGG 163
:|||||:|||||:||||| ||| |||||||||:|||||
490 erArgSerProGlnArgArgGlyArgSerArgSerProGlnArgProGly 506
164 TGGTCCGCTTCCAAAGAAG...CAGCCAATAAGCAGAGCAAGCACTCGAC 210
|||||:|||||:||||| ||| |||||||||:|||||
507 TrpSerArgSerArgAsnThrGlnArgArgGlyArgSerArgSerAlaAr 523
211 GAGATCTCGACGAATATTCGTCAGCGCGCGGTCCAAATCTCGAGGCGCGA 260
|||||:|||||:||||| ||| |||||||||:|||||
523 gArgGlyArgSerHisSerArgSerProAlaThrArgGlyArgSerArgS 540
261 CGAGGAGCAGCAGCAGCGCG 280
540 erArgThrProAlaArgArg 546
```


OM of: US-09-462-480-3 to: A_Geneseq_032802:* out_format : pfs
Date: Jul 22, 2002 1:22 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame2.n2p.model -DEV=xih
-O=/cgn2_1/USPTO.spool/US09462480/runat_18072002_164418_19413/app_query.fasta_1.2850
-DB-A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.500
-XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40 cdi -LISP=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480 -CGN1_1_57 -NCPU=6 -TCPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-462-480-3
Query length: 481
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 134.850000

score_list:

Sequence	Strd Orig	zScore	Escore	Len	! Documentation
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW32452 +	...	713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW32384 +	...	713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW61705 +	...	713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW64338 +	...	713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW39135 +	...	713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW38992 +	...	713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAW19844 +	...	713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW27229 +	...	713.00	1059.23	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW21946 +	...	713.00	1059.23	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW27228 +	...	373.00	559.56	3.7e-23	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW21945 +	...	373.00	559.56	3.7e-23	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW54248 +	...	147.50	215.51	0.0002	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW54248 +	...	145.50	216.42	0.0003	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAW79841 +	...	141.50	195.79	0.0009	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW24091 +	...	140.00	200.22	0.0010	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG21932 +	...	136.50	195.59	0.0020	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW79128 +	...	134.00	196.80	0.0031	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW31855 +	...	134.00	190.16	0.0034	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW31852 +	...	134.00	187.68	0.0035	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG13470 +	...	133.50	198.23	0.0033	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW20852 +	...	132.50	196.94	0.0040	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAW39138 +	...	132.50	185.23	0.0046	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAW39141 +	...	131.00	183.85	0.0061	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW72204 +	...	128.50	167.31	0.0116	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG21919 +	...	127.50	185.32	0.0110	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG30150 +	...	127.50	183.22	0.0113	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAW31516 +	...	127.50	169.28	0.0134	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW12000 +	...	127.50	169.26	0.0134	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW18664 +	...	125.50	180.85	0.0164	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG03626 +	...	125.50	184.17	0.0172	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABW70063 +	...	124.50	178.11	0.0202	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABW11016 +	...	124.00	175.68	0.0227	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG18528 +	...	123.00	180.33	0.0255	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABW65293 +	...	122.50	166.64	0.0329	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG04558 +	...	121.00	176.96	0.0376	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG19764 +	...	121.00	175.80	0.0381	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW68408 +	...	121.00	173.73	0.0391	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1992.DAT:AAW28150 +	...	121.00	173.02	0.0395	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW28595 +	...	121.00	169.39	0.0413	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG20365 +	...	121.00	165.11	0.0435	

/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW13464 - 120.50 163.25 0.04
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW76734 - 120.00 162.05 0.05
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW17737 - 119.50 175.43 0.05
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW17736 - 119.50 174.53 0.05
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW17735 - 119.50 171.28 0.05

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW32452

seq_documentation_block:
ID AAW32452 standard; Protein; 368 AA.

XX AAW32452;

DT 09-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb37-FL.

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers
FT Misc-difference 10 /note= "Any amino acid"

XX WO9709428-A2.
XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US14674.
PR 12-JUL-1996; 96US-0680574.
PR 01-SEP-1995; 95US-0523436.
PR 22-SEP-1995; 95US-0533634.
PR 22-MAR-1996; 96US-0620874.
PR 05-JUN-1996; 96US-0659683.

PA (CORI-) CORIXA CORP.

PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
PI Twardzik DR, Vedvick TH;
WPI; 1997-192903/17.
New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
useful in vaccines for prevention or treatment of tuberculosis, also
for diagnosis

Example 3; Page 146-147; 168pp; English.

CC A new immunogenic polypeptide has been developed comprising an
immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
its variant differing only in conservative substitutions and/or
modifications). The present sequence represents a M.tuberculosis
antigen, Tb37-FL. The immunogenic protein, and fusion proteins
containing one or more of the proteins or one of the proteins plus
ESAN-6, are useful in vaccines preferably when formulated with a
non-specific adjuvant, to induce an immune response against
M.tuberculosis (for treatment or prevention).

Sequence 368 AA;

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x AAW32452 ..
Align seg 1/1 to: AAW32452 from: 1 to: 368


```

AC AAW81705;
XX
XX 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide Tb37-FL.
XX
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX vaccine; pharmaceutical; infection; diagnosis.
XX
XX Mycobacterium tuberculosis.
XX
XX Key Location/Qualifiers
XX Misc-difference 10
XX /label= unknown
XX
XX WO9816646-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18293.
XX
XX 13-MAR-1997; 97US-0818112.
XX 11-OCT-1996; 96US-0730510.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX WPI; 1998-261042/23.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
XX to develop products for the detection of M. tuberculosis infection
XX and for diagnosis, treatment and prevention of tuberculosis
XX
XX Example 3B; Page 137-138; 230pp; English.
XX
XX This sequence represents an immunogenic portion of a soluble
XX Mycobacterium tuberculosis (MT) antigen which can be used in a method
XX for inducing protective immunity against tuberculosis (TB). This
XX sequence can be formulated into vaccines and/or pharmaceutical
XX compositions for immunising against M. tuberculosis infection or may
XX be used for the diagnosis of tuberculosis.
XX
XX Sequence 368 AA;

alignment_scores:
  Quality: 713.00 Length: 143
  Ratio: 5.057 Gaps: 0
  Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x AAW81705 ..

Align seg 1/1 to: AAW81705 from: 1 to: 368

1 CTGCAGCAGTGACGCTGTTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242

51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGGCTGCTCGGCACCATC 100
|||||
242 YASNProAlaAspGluAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGTGTCGACACCATCCGCTGGCTGGTGATGATGAGCCCGCCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275

151 GGCTGCTGCGCGGAGTGCTACCTGGCGCAGGTGGGTGCTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292

```

```

201 CACGCGGCTGATGCTCAGCTGATCGAAAAAGCCGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309

251 TGCCGGGCGCTGTTGCGGATCCTCGGTGACGGTGGCGCGCTCGCGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325

301 GGTCGGGAGCGATGGCCAGGTTCGCAATCGCGGGCTCCACACGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342

351 GGCTGCTGTCGCGCGCCGCGCTCGCGCAGGAGCGGTGAAGAAGACGAC 400
|||||
342 OGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359

401 AGGACGACTGGGACGAGAGGAGGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64338

seq_documentation_block:
ID AAW64338 standard; Protein; 368 AA.
XX
XX AAW64338;
XX
XX 09-NOV-1998 (first entry)
XX
XX Mycobacterium tuberculosis antigen Tb37-FL.
XX
XX Tuberculosis; infection; diagnosis; antigen; Tb37-FL.
XX
XX Mycobacterium tuberculosis strain H37Rv.
XX
XX Key Location/Qualifiers
XX Misc-difference 10
XX /note= "unidentified"
XX
XX WO9816645-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18214.
XX
XX 13-MAR-1997; 97US-0818111.
XX 11-OCT-1996; 96US-0729622.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX WPI; 1998-251292/22.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX to develop products for the detection of M. tuberculosis infection
XX and diagnosis of tuberculosis
XX
XX Example 3; Page 143-144; 250pp; English.
XX
XX This polypeptide comprises Mycobacterium tuberculosis antigen
XX Tb37-FL. It is encoded by genomic DNA isolated from a M.
XX tuberculosis strain H37Rv genomic library using a probe from
XX clone Tb38-1 (see AAW44384). The invention relates to compositions
XX and methods for diagnosing tuberculosis. It provides polypeptides
XX (see AAW64291-W64379) comprising an antigenic portion of a soluble
XX M. tuberculosis antigen, or an immunogenic portion of an M.
XX tuberculosis antigen, as well as DNA sequences encoding such
XX polypeptides, recombinant expression vectors and transformed or
XX transfectected host cells. Also claimed are methods and diagnostic
XX kits for detecting M. tuberculosis infection in a patient using
XX these polypeptides, antibodies or oligonucleotide probes and

```

CC primers, for the diagnosis of tuberculosis.

XX Sequence 368 AA;

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-3 x AAW64338 ..

Align seg 1/1 to: AAW64338 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTCGTTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAAGCCGCGAGATGGGCTGCTCGGCACCAATC 100
242 YAsnProAlaAspGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGGAACCATCCGCTGCTGGTGGATCAGGCCGCCAGCGCGCGG 150
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGCGGAGTGCCTACCTGGCGCAGGTGGTGGTACCCG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTGATCGAAGCCGGTTGCCCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309
251 TCCCGCGGCTGTTGCGGATCGTCGGTGACGGGTGGCGGCTCGGTG 300
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCTGTCGGGAGTGGCCAGGCGTTCGCAATCCGCGCGCTCCACGCGC 350
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.AAY39135

seq_documentation_block:

ID AAY39135 standard; Protein; 368 AA.

XX AC AAY39135;

XX DT 05-NOV-1999 (first entry)

XX DE M. tuberculosis antigen Tb37-FL amino acid sequence.

XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.

XX OS Mycobacterium tuberculosis.

XX PN W09942076-A2.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US03268.

XX

PR 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;
XX WPI; 1999-527409/44.

XX New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX Example 3; Page 132-133; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.

XX Sequence 368 AA;

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-3 x AAY39135 ..

Align seg 1/1 to: AAY39135 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTCGTTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAAGCCGCGAGATGGGCTGCTCGGCACCAATC 100
242 YAsnProAlaAspGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGGAACCATCCGCTGCTGGTGGATCAGGCCGCCAGCGCGCGG 150
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGCGGAGTGCCTACCTGGCGCAGGTGGTGGTACCCG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTGATCGAAGCCGGTTGCCCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309
251 TCCCGCGGCTGTTGCGGATCGTCGGTGACGGGTGGCGGCTCGGTG 300
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCTGTCGGGAGTGGCCAGGCGTTCGCAATCCGCGCGCTCCACGCGC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrAr 342
351 GGTCTGTCGGCGCGGACGCTCGCGCAGGCGGTCAAGAACACGACG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAsp 359
```

401 AGGACGACTGGGACGACGAGGACGACTGG 429
 |||||
 359 luAspAspTirpAspGluAspAspTrp 368

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AA138992

seq_documentation_block:

ID AAY38992 standard; Protein: 368 AA.

AC AAY38992;

DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen protein Tb37-FL.

KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 vaccine; immunity.

OS Mycobacterium tuberculosis.

PN WO9942118-A2.

PD 26-AUG-1999.

PF 17-FEB-1999; 99WO-US03265.

PR 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

PA (CORI-) CORIXA CORP.

XX

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;

XX

XX WPI; 1999-527416/44.

XX

PT New polypeptide comprising antigenic portions of M. tuberculosis

PS Example 3; Page 177-179; 323pp; English.

XX

CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.

XX

XX Sequence 368 AA;

alignment_scores:

Quality: 713.00 Length: 143
 Ratio: 5.057 Gaps: 0
 Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-3 x AAY38992 ..

Align seg 1/1 to: AAY38992 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGTTGTCAGCCAGGTGGCGCGCAGCGGCGCG 50

|||||
 226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG1 242

51 CAACCCAGCCGACGAGGAGCGCGCAGATGGGCTGCTGCGCAGCAGTC 100

|||||
 242 YAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTCGAACATCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 150

|||||
 259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCCTGCTGCGCGGAGTCGCTACCTGCGCGCAGGTGGTGGTTCGTTGACCCG 200
 |||||
 276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
 201 CACGCCGTGATGTCACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 250
 |||||
 292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
 251 TGCGCGGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 300
 |||||
 309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
 301 GGTCCGGAGCGATGGCGCAGGTTTCGCAATCCGCGCGCTCCACACGCCC 350
 |||||
 356 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
 351 GGTCTGTCG 400
 |||||
 342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
 401 AGGACGACTGGGACGACGAGGACGACTGG 429
 |||||
 359 luAspAspTirpAspGluAspAspTrp 368

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT: AAB19844

seq_documentation_block:

ID AAB19844 standard; Protein: 368 AA.

XX AAB19844;

XX

DT 05-MAR-2001 (first entry)

DE Mycobacterium tuberculosis protein MTBN3.

XX

KW MTBN3; tuberculosis; BCG; vaccine; infection; diagnosis.

XX

OS Mycobacterium tuberculosis.

XX

PN WO200066157-A1.

XX

PD 09-NOV-2000.

XX

PF 04-MAY-2000; 2000WO-US12257.

XX

PR 04-MAY-1999; 99US-0132505.

XX

PA (PUBL-) PUBLIC HEALTH RES INST NEW YORK.

XX

PI Gennaro ML;

XX

DR WPI; 2001-007153/01.

DR N-PSDB; AAA89037.

XX

PT Novel polypeptide encoded by open reading frames present in
 PT Mycobacterium tuberculosis genome and not by the BCG strain of M.
 PT bovis, useful as vaccine and for diagnosing tuberculosis infection -
 PT
 PS Claim 11; Fig 1; 35pp; English.
 XX
 CC The present sequence is that of the Mycobacterium tuberculosis
 CC MTBN3 protein. This is 1 of 8 proteins, i.e. MTBN1-8 (see
 CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)
 CC identified as being present in the genome of M. tuberculosis but
 CC absent from the genome of the BCG strain of Mycobacterium bovis.
 CC MTBN1-8 represent reagents that are useful in discriminating between
 CC M. tuberculosis and BCG and, in particular, for diagnostic methods
 CC which discriminate between exposure of a subject to M. tuberculosis
 CC and vaccination with BCG. The invention features these MTBN
 CC polypeptides, functional fragments of them, DNA encoding them,
 CC vectors, transformed cells, and diagnostic, therapeutic, and
 CC prophylactic (vaccine) methods, including genetic vaccination
 CC methods.

XX
SQ Sequence 368 AA;

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-3 x AAB19844

Align seg 1/1 to: AAB19844 from: 1 to: 368

```

1 CTGCAGCAGGTGACGCTCGTTGTTACGACGAGTGGCGGCACCGGGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyGly 242
51 CAACCCAGCCGACGAGGAGGAGCGCGAGATGGGCTGTCGGCACCATC 100
|||||
242 YAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGCTGAACCATCCGCTGGCTGTGATCAGCCGCCAGCGCGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGGAGTCGCTACTGCGCGAGGTGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CAGCGCGTGTATGCTCAGCTGATCGAAGCCGGTTCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTTCCCGGATCGTGGTGACGGTGCGCGCTCCGGTG 300
|||||
309 eProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGAGGAGGATGGCGAGGTTTCGCAATCCGCGGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTGTGTGCGCGCGGACCGCTCGCGCAGGAGCGGTGAAGAAGACGAC 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAsp 359
401 AGGACGACTGGGACGAGGAGGAGGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

```

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72929

seq_documentation_block:

ID_AAW72929 standard; Protein; 371 AA.

XX

AC AAW72929;

XX 21-JAN-1999 (first entry)

DE Mycobacterium tuberculosis antigen RD1-ORF5.

XX Mycobacterium tuberculosis; antigen; vaccine; immunological;
KW Immunogen; infection.

XX Mycobacterium tuberculosis.

OS WO9844119-A1.

XX 08-OCT-1998.

XX 01-APR-1998; 98WO-DK00132.

XX 05-JAN-1998; 98US-0070488.

PR 02-APR-1997; 97DK-0000376.
PR 18-APR-1997; 97US-0044624.
PR 10-NOV-1997; 97DK-0001277.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB;
PI Rosenkrands I, Weidinger K;

XX WPI: 1998-542705/46.

DR N-PSDB; AAV63939.

XX New isolated mycobacteria polypeptides and nucleic acids - used for
PT developing products for the diagnosis of or vaccination against
PT mycobacterial infections, particularly tuberculosis

XX Claim 1: Page 200-202; 163pp; English.

XX The present sequence represents a Mycobacterium tuberculosis protein.
CC Products from the present invention, which describes protein fragments
CC and nucleic acid fragments derived from M.tuberculosis, can be used in
CC the detection of and prevention of mycobacterial infections. In
CC particular, the proteins and nucleic acids can be used for the diagnosis
CC of or vaccination against tuberculosis caused by M. tuberculosis,
CC M. africanum or M. bovis.

XX SQ Sequence 371 AA;

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-3 x AAW72929

Align seg 1/1 to: AAW72929 from: 1 to: 371

```

1 CTGCAGCAGGTGACGCTCGTTGTTACGACGAGTGGCGGCACCGGGCGG 50
|||||
229 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyGly 245
51 CAACCCAGCCGACGAGGAGGAGCGCGAGATGGGCTGTCGGCACCATC 100
|||||
245 YAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 262
101 CGCTGCTGAACCATCCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 150
|||||
262 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 278
151 GGCCTGCTGCGCGGAGTCGCTACTGCGCGAGGTGGTGGTGGTGGTGGTGG 200
|||||
279 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 295
201 CAGCGCGTGTATGCTCAGCTGATCGAAGCCGGTTCGCCCTCGGTGA 250
|||||
295 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 312
251 TGCCGGCGGCTGTTCCCGGATCGTGGTGACGGTGCGCGCTCCGGTG 300
|||||
312 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 328
301 GGTCCGGAGGAGGATGGCGAGGTTTCGCAATCCGCGGCTCCACGAGCC 350
|||||
329 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 345
351 GGGTGTGTGCGCGCGGACCGCTCGCGCAGGAGCGGTGAAGAAGACGAC 400
|||||
345 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAsp 362
401 AGGACGACTGGGACGAGGAGGAGGACTGG 429
|||||

```

362 luaspaspTrpaspGluGluAspAspTrp 371

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AA1999

seq_documentation_block:

ID AAY21946 standard; Protein; 371 AA.

AC AAY21946;

DT 06-SEP-1999 (first entry)

DE Amino acid sequence of antigen RD1-ORF5.

XX Immunogenic; Mycobacterium tuberculosis; immune response; infection; tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB; pharmacological; vaccination; M. africanum; M. bovis; CFP7A; CFP30A; CFP7B; CFP19; CFP27; CFP30A; RD1-ORF; CFP10A; CFP16; CFP19; CFP23; CFP25A; CFP30B; CFP7B.

XX Mycobacterium tuberculosis.

OS WO9924577-A1.

PN 20-MAY-1999.

PD 08-OCT-1998; 98WO-DK00438.

PF 01-APR-1998; 98WO-DK00132.

PR 10-NOV-1997; 97DK-0001277.

PR 05-JAN-1998; 98US-0070488.

XX (STAT-) STATENS SERUM INST.

PA Andersen P, Skjot R;

XX WPI; 1999-347282/29.

DR N-PSDB; AAX81046.

XX New immunogenic fragment of Mycobacterium tuberculosis.

PT Example 2; Page 219-220; 265pp; English.

XX The invention describes a substantially pure immunogenic polypeptide fragment (I) from Mycobacterium tuberculosis that is able to evoke a protective immune response against infections by mycobacteria belonging to the tuberculosis complex. The invention provides a (1) fusion polypeptide comprising at least one polypeptide fragment (I) and at least one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second different amino acid sequence from M. tuberculosis, and/or including a sequence which protects the first amino acid sequence from in vivo degradation or post-translational processing; (3) a nucleic acid fragment that encodes the above polypeptides. The polypeptides and nucleic acid are useful as pharmaceuticals, for diagnosis of and as antigens for vaccination against TB caused by Mycobacterium tuberculosis, africanum or bovis. The polypeptides are also useful for diagnosing ongoing or previous sensitization in an animal with bacteria belonging to the tuberculosis complex. The invention also describes the use of CFP7A or CFP30A or a T-cell epitope of for the induction of a strong immune response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell epitope of for diagnosis of TB in a mammal by performing a DTH type skin test; use of CFP27, CFP30A, RD1-ORF2, RD1-ORF3, RD1-ORF5, MPT59-ESAT6, ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-cell epitope of for the preparation of an immunological composition; and for the preparation of a subunit vaccine.

XX Sequence 371 AA;

alignment_scores:

Quality:	713.00	Length:	143
Ratio:	5.057	Gaps:	0
Percent Similarity:	98.601	Percent Identity:	96.503

alignment_block:

US-09-462-480-3 x AAY21946

Align seg 1/1 to: AAY21946 from: 1 to: 371

```

1 CTGCAGCAGGTGACGTCGTGTTTCAGCCAGGTGGGGCGGCACCGCGGGCGG 50
|||||
229 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 245
51 CAACCCAGCCGAGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 100
|||||
245 yasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 262
101 CGCTGTCGAACCATCGCTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 150
|||||
262 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 278
151 GGCCTGCTGCGCGGCGGAGTCGCTACCTGGCGGCGGAGTGGTGGTGGT 200
|||||
279 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 295
201 CAGCGCGCTGATGCTCAGCTGATCGAAAGCGGTTGCCCGCTCGGTGA 250
|||||
295 gThrProLeuMetSerGlnLeuIleGlnLysProValAlaProSerVal 312
251 TCGCGGCGGCTGTCGCGGATCGTCGGTGACGGTGGCGGCGGCGGCGG 300
|||||
312 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 328
301 GGTCCGGGAGCGATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 350
|||||
329 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 345
351 GGGTCTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 400
|||||
345 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAsp 362
401 AGGACGACTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 429
362 luaspaspTrpaspGluGluAspAspTrp 371

```

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72928

seq_documentation_block:

ID AAW72928 standard; Protein; 139 AA.

XX AC AAW72928;

XX DT 21-JAN-1999 (first entry)

XX DE Mycobacterium tuberculosis antigen RD1-ORF4.

XX KW Mycobacterium tuberculosis; antigen; vaccine; immunological; immunogen; infection.

XX OS Mycobacterium tuberculosis.

XX PN WO9844119-A1.

XX PD 08-OCT-1998.

XX PF 01-APR-1998; 98WO-DK00132.

XX PR 05-JAN-1998; 98US-0070488.

PR 02-APR-1997; 97DK-0000376.

PR 18-APR-1997; 97US-0044624.

PR 10-NOV-1997; 97DK-0001277.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB;

PI Rosenkrands I, Weidigh K;

```
XX WPI; 1998-542705/46.
DR N-PSDB; AAV63938.
XX
PT New isolated mycobacteria polypeptides and nucleic acids - used for
PT developing products for the diagnosis of or vaccination against
PT mycobacterial infections, particularly tuberculosis
XX
XX Claim 1; Page 198; 163pp; English.
XX
XX The present sequence represents a Mycobacterium tuberculosis protein.
CC Products from the present invention, which describes protein fragments
CC and nucleic acid fragments derived from M.tuberculosis, can be used in
CC the detection of and prevention of mycobacterial infections. In
CC particular, the proteins and nucleic acids can be used for the diagnosis
CC of or vaccination against tuberculosis caused by M. tuberculosis,
CC M. africanum or M. bovis.
XX
XX Sequence 139 AA;
SQ
alignment_scores:
  Quality: 373.00      Length: 67
  Ratio: 5.567        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.507
alignment_block:
US-09-462-480-3/rev x AAW72928
Align seg 1/1 to: AAW72928 from: 1 to: 139
203 GTCCGGGTCAACGACCCACTGCGCAGGTAGCGACTCCGCGCGCAGCAG 154
  ::::|||||
1 MetArgValAsnAspProAlaProGlySerAspSerAlaArgSerAr 17
153 GCCCGCGCGCGCTGGCGCTGATCCAGCCAGCCAGCGGATGTTGCACA 104
  :|||||
17 gProAlaProAlaLeuGlyProAspProAlaSerGlyTrpPheAsps 34
103 GCGGACTGTGTCGAGCAGCCCATCTGCGCGCTTCCTCGTGGCTGGG 54
  :|||||
34 erGlyLeuValProSerArgProIleCysAlaAlaSerSerAlaGly 50
53 TTCCCGCGCGGTGCGCGCCACTGGCTGGAACACAGCGTACCTGCTG 4
  :|||||
51 LeuProProValProThrTrpLeuAsnAsnAspValThrCysCys 67
3 C 3
67 s 67
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV21945
seq_documentation_block:
ID AAV21945 standard; Protein; 139 AA.
XX
AC AAV21945;
XX
DT 06-SEP-1999 (first entry)
XX
DE Amino acid sequence of antigen RDI-ORF4.
XX
KW Immunogenic; Mycobacterium tuberculosis; immune response; infection;
KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;
KW pharmaceutical; vaccination; M. africanum; M. bovis; CFP7A; CFP30A;
KW CFP7B; CFP19; CFP27; CFP30A; RDI-ORF; CFP10A; CFP16; CFP19; CFP23;
KW CFP25A; CFP30B; CFP7B.
XX
OS Mycobacterium tuberculosis.
XX
PN W09924577-A1.
XX
PD 20-MAY-1999.
```

```
XX 08-OCT-1998; 98WO-DK00438.
PF
XX
XX 01-APR-1998; 98WO-DK00132.
PR
XX 10-NOV-1997; 97DK-0001277.
PR
XX 05-JAN-1998; 98US-0070488.
PR
XX (STAT-) STATENS SERUM INST.
PA
XX
XX Andersen P, Skjot R;
PI
XX
XX WPI; 1999-347282/29.
DR
XX N-PSDB; AAX81045.
DR
XX
XX New immunogenic fragment of Mycobacterium tuberculosis
PT
XX
XX Example 2; Page 216; 265pp; English.
PS
XX
CC The invention describes a substantially pure immunogenic polypeptide
CC fragment (I) from Mycobacterium tuberculosis that is able to evoke a
CC protective immune response against infections by mycobacteria belonging
CC to the tuberculosis complex. The invention provides a (1) fusion
CC polypeptide comprising at least one polypeptide fragment (I) and at least
CC one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell
CC epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second
CC different amino acid sequence from M. tuberculosis, and/or including a
CC sequence which protects the first amino acid sequence from in vivo
CC degradation or post-translational processing; (3) a nucleic acid fragment
CC that encodes the above polypeptides. The polypeptides and nucleic acid
CC are useful as pharmaceuticals, for diagnosis of and as antigens for
CC vaccination against TB caused by Mycobacterium tuberculosis, africanum or
CC bovis. The polypeptides are also useful for diagnosing ongoing or
CC previous sensitization in an animal with bacteria belonging to the
CC tuberculosis complex. The invention also describes the use of CFP7A or
CC CFP30A or a T-cell epitope of for the induction of a strong immune
CC response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell
CC epitope for diagnosis of TB in a mammal by performing a DTH type skin
CC test; use of CFP27, CFP30A, RDI-ORF2, RDI-ORF3, RDI-ORF5, MPT59-ESAT6,
CC ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-
CC cell epitope of for the preparation of an immunological composition; and
CC for the preparation of a subunit vaccine.
XX
XX Sequence 139 AA;
SQ
```

```
alignment_scores:
  Quality: 373.00      Length: 67
  Ratio: 5.567        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.507
alignment_block:
US-09-462-480-3/rev x AAY21945
Align seg 1/1 to: AAY21945 from: 1 to: 139
203 GTCCGGGTCAACGACCCACTGCGCAGGTAGCGACTCCGCGCGCAGCAG 154
  ::::|||||
1 MetArgValAsnAspProAlaProGlySerAspSerAlaArgSerAr 17
153 GCCCGCGCGCGCTGGCGCTGATCCAGCCAGCCAGCGGATGTTGCACA 104
  :|||||
17 gProAlaProAlaLeuGlyProAspProAlaSerGlyTrpPheAsps 34
103 GCGGACTGTGTCGAGCAGCCCATCTGCGCGCTTCCTCGTGGCTGGG 54
  :|||||
34 erGlyLeuValProSerArgProIleCysAlaAlaSerSerAlaGly 50
53 TTCCCGCGCGGTGCGCGCCACTGGCTGGAACACAGCGTACCTGCTG 4
  :|||||
51 LeuProProValProThrTrpLeuAsnAsnAspValThrCysCys 67
3 C 3
67 s 67
```


67 s 67

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAE01594

seq_documentation_block:

ID AAE01594 standard; Protein; 314 AA.

AC AAE01594;

XX 17-JUL-2001 (first entry)

XX Human gene 12 encoded secreted protein HHSGW69, SEQ ID NO:144.

XX Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiotensin II; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulvar;
KW cell culture; chemotaxis; food additive; chromosome 17;
KW binding partner identification; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1..26

XX FT /label= Signal_peptide

XX FT 27..55

XX FT /label= Mature_human_secreted_protein

XX FT Misc-difference 129

XX FT /label= Unknown

XX FT /note= "Encoded by MCG"

XX PN W0200134623-AA1.

XX PD 17-MAY-2001.

XX PF 01-NOV-2000; 2000WO-US30037.

XX PR 05-NOV-1999; 99US-0163577.

XX PR 30-JUN-2000; 2000US-0215137.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Komatsoulis GA, Moore PA;

XX XX WPI; 2001-316490/33.

XX DR N-PSDB; AAD05437.

XX Nucleic acids encoding 29 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -

XX Claim 11; Page 496-497; 535pp; English.

XX AAD05389-AAD05473 represent cDNAs corresponding to 29 human secreted
CC protein genes, and AAE01546-AAE01630 represent the proteins they encode.
CC AAE01631-AAE01660 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 29 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,

CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiotensin II, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.

XX Sequence 314 AA;

alignment_scores:

Quality: 147.50 Length: 163

Ratio: 2.049 Gaps: 10

Percent Similarity: 44.172 Percent Identity: 33.742

alignment_block:

US-09-462-480-3/rev x AAE01594

Align seg 1/1 to: AAE01594 from: 1 to: 314

472 CGCGCGCGGTGGCGGGGAAGTCTGTGTCTACGGAGCTCACCAGTCG 423

|||||

126 ArgProGly***Pro..... 130

422 TCCTCTTCGTCCTCCAGTCGT.....CCTC 400

|||||

131 .ProLeuProProSerSerSerGlyAsnAspGluThrProArgLeuProG 147

399 GTCGCTCTTTCACGCTCCTCGCGAGCGGTGCGCGCGACGACCG 350

|||||

147 InArgAsnLeu...SerLeuSerSerThrProLeuProSerPro 162

349 GGCTGGTGGAGCGCGGATTCGGAACCTCGCCCATCGTCCCGGACCC 300

|||||

163 Gly.....ArgSerGlyPro...LeuProProPr 171

299 ACCGAGGCGCGCCACCGTCACGACGATCGCGCAACAGCGCGCGCAT 250

|||||

171 OProSerGluArg...ProProProProValArgAspProProGlyArgS 187

249 CA.....CCGAGGGGGCAACCGGCTTTTCGATCAGCTGAGACATCA 209

|||||

187 erGlyProLeuProProProProProValSerArgAsnGlySerThrSer 203

208 GCGGCGTGGGTCAACGACCCACCTGCGCCA.....GGTAGCGAC 168

|||||

204 ArgAlaLeuProAlaThr..ProGlnLeuProSerArgSerGlyValAsp 219

167 TCCGCGCGCAGCAGCGCGCGCTGGGCGCTGATCCACAGCCAGCCAG 118

|||||

220 SerProArgSerGlyProArgProProLeuProProAspArgPro..... 234

117 CGGATGTTTCGACAGCGGACTGGTCCGAGCAGGCGCCCATCTCCGCGGTT 68

234

67 CTCTGTCGGTGGTGGTGGCGCGCGCGGTGCGCGCCACC 30

|||||

235SerAlaGlyAlaProProProProProSer 245

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AAE015428

seq_documentation_block:

ID AAE015428 standard; Protein; 200 AA.

XX

STTBTB4 TACT 00T

[illegible]

2

alignment_scores:	
Quality:	140.00
Ratio:	2.000
Percent Similarity:	42.945
Percent Identity:	32.515
Gaps:	9
Length:	163

```
alignment_block:
US-09-462-480-3/rev x AAY24091
Align seg 1/1 to: AAY24091 from: 1 to: 504

472 CGGCGGGGTGGCGGGAAGTCTGTGTCATTACGGGAGCTCACCAGTGG 423
|||||
315 ArgProGlyProPro..... 319
422 TCCTCTTCGTCCAGTCGT.....CCTC 400
|||||
320 .ProLeuProProSerSerGlyAsnAspGluThrProArgLeuProG 336
|||||
399 GTCGCTTCTTCACGCTCTGCGGAGCGGTGCGCGCGACGACGACCG 350
::||| ||| ::| ::| ::| ::| ::| ::| ::| ::|
336 InArgAsnLeu...SerLeuSerSerThrProProLeuProSerPro 351
|||
349 GCGTGTGAGCGCGCGGATTGCGAACCTGGCCCATGCTCCCGGACCC 300
||| ::| ::| ::| ::| ::| ::|
352 Gly.....ArgSerGlyPro...LeuProProPr 360
299 ACCGGAGCGCGCCACCGCTCACCGACGATCCGCAACAGCGCGGCAT 250
||| ::| ::| ::| ::| ::| ::|
360 oValProSerGluArgProProProValArgAspProProGlyArgS 377
249 CA.....CCGAGGGGGCAACCGGCTTTTCGATCAGCTGAGACATCA 209
||| ||| ::| ::| ::| ::| ::| ::|
377 erGlyProLeuProProProProValSerArgAsnGlySerThrSer 393
208 GCGGGGTGGGTCAACGACCCACCTGCGCCA.....GGTAGCGAC 168
||| ::| ::| ||| ||| ||| |||
394 ArgAlaLeuProAlaThr..ProGlnLeuProSerArgSerGlyValAsp 409
167 TCCGCGCGCAGCAGCGCGCGCGCTGGGGCCTGATCCACAGCCAG 118
||| ||| ||| ||| ||| ||| ||| |||
410 SerProArgSerGlyProArgProProLeuProProAspArgPro.... 424
117 CGGATGGTTCGACAGCGGACTGGTCCGAGCAGGCCCATCTGCGGGCTT 68
424 ..... 424
67 CCTCGTGGGTGGTGGCGCGCGGTGGCGGCCACC 30
||||| ||| ||| ||| ||| |||
425 .....SerAlaGlyAlaProProProProSer 435
```

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-08-818-112-111 +	713.00	1066.91	368 !	1.7e-52
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-08-818-111-109 +	713.00	1066.91	368 !	1.7e-52
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-09-056-556-111 +	713.00	1066.91	368 !	1.7e-52
/cgn2_6/ptodata/2/iaa/5B COMB	pep:US-09-010-928B-28 +	128.00	176.33	745 !	0.0026
/cgn2_6/ptodata/2/iaa/5B COMB	pep:US-09-010-928B-2 +	128.00	176.91	870 !	0.0026
/cgn2_6/ptodata/2/iaa/5B COMB	pep:US-08-726-366A-56 +	125.50	187.54	181 !	0.0032
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-09-413-814-78 -	125.00	172.26	882 !	0.0047
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-09-232-468A-8 -	121.00	173.39	404 !	0.0089
/cgn2_6/ptodata/2/iaa/5B COMB	pep:US-09-080-897-2 -	120.50	162.30	1248 !	0.0119
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-09-323-735-2 -	120.50	162.30	1248 !	0.0119
/cgn2_6/ptodata/2/iaa/6A COMB	pep:US-08-899-595-3 -	120.00	161.06	1315 !	0.0133
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-08-483-533-41 -	118.50	170.80	355 !	0.0141
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-09-283-471A-41 -	118.50	170.80	355 !	0.0141
/cgn2_6/ptodata/2/iaa/PTCUS COMB	pep:PCR-US91-06532-3	118.50	170.80	355 !	0.0141
/cgn2_6/ptodata/2/iaa/5B COMB	pep:US-09-010-928B-4 +	118.00	161.45	907 !	0.0183
/cgn2_6/ptodata/2/iaa/5A COMB	pep:US-07-945-288-2 +	117.50	153.64	1958 !	0.0231
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-08-483-533-40 -	116.00	169.75	264 !	0.0217
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-09-283-471A-40 -	116.00	169.75	264 !	0.0217
/cgn2_6/ptodata/2/iaa/5A COMB	pep:US-07-928-611-22 -	112.00	163.62	267 !	0.0472
/cgn2_6/ptodata/2/iaa/5B COMB	pep:US-08-487-811A-13 -	112.00	163.62	267 !	0.0472
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-09-060-694-13 -	112.00	163.62	267 !	0.0472
/cgn2_6/ptodata/2/iaa/PTCUS COMB	pep:PCR-US93-07370-13 -	112.00	163.62	267 !	0.0472
/cgn2_6/ptodata/2/iaa/5A COMB	pep:US-08-056-051-6 -	112.00	158.49	467 !	0.0520
/cgn2_6/ptodata/2/iaa/5A COMB	pep:US-07-928-611-22 -	112.00	158.49	467 !	0.0520
/cgn2_6/ptodata/2/iaa/5B COMB	pep:US-08-487-811A-22 -	112.00	158.49	467 !	0.0520
/cgn2_6/ptodata/2/iaa/PTCUS COMB	pep:US-09-060-694-22 -	112.00	158.49	467 !	0.0520
/cgn2_6/ptodata/2/iaa/PTCUS COMB	pep:PCR-US93-07370-22 -	111.50	158.49	467 !	0.0520
/cgn2_6/ptodata/2/iaa/PTCUS COMB	pep:PCR-US91-06532-2 -	111.50	163.00	263 !	0.0518
/cgn2_6/ptodata/2/iaa/backflesl pep	pep:5168049-3 -	110.00	164.06	183 !	0.0650
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-08-483-533-26 -	110.00	160.91	258 !	0.0690
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-09-283-471A-26 -	110.00	160.91	258 !	0.0690
/cgn2_6/ptodata/2/iaa/5A COMB	pep:US-08-131-365B-38 -	110.00	147.24	1147 !	0.0897
/cgn2_6/ptodata/2/iaa/5B COMB	pep:US-08-668-123-38 -	110.00	147.24	1147 !	0.0897
/cgn2_6/ptodata/2/iaa/backflesl pep	pep:5171843-7 +	109.50	156.66	378 !	0.0813
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-09-199-637A-351 -	109.00	154.37	447 !	0.0922
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-09-199-637A-277 +	109.00	152.33	558 !	0.0959
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-09-041-886-23 -	109.00	145.43	1185 !	0.1094
/cgn2_6/ptodata/2/iaa/5A COMB	pep:US-07-945-283-2 -	109.00	140.83	1958 !	0.1195
/cgn2_6/ptodata/2/iaa/backflesl pep	pep:5171843-7 +	108.50	159.63	232 !	0.0905
/cgn2_6/ptodata/2/iaa/6B COMB	pep:US-08-556-978B-59 -	108.50	152.72	493 !	0.033

```
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTCGCGCGGAGTCTGCTACCTGGCGCAGGTGGTCTTGCACCG 200
|||||
276 GlyLeuLeuArgAlaGlySerLeuProGlyAlaGlySerLeuThrAr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCGGTTCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGlyProValAlaProSerValM 309
251 TCCCGGCGGCTGTTCCCGGATCGCTGAGCGGTGGCGGCTCCGCTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGCGGATCGCGGAGGTTCGCAATCCCGCGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTGCGCGCGGACCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGGAGGAGGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-818-111-109

```
seq_documentation_block:
; Sequence 109, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-818-111-109
```

```
alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x US-08-818-111-109 ..
Align seg 1/1 to: US-08-818-111-109 from: 1 to: 368
```

```
1 CTGACGAGGTGACGTCTGTTTTCAGCCAGGTGGCGCGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
|||||
51 CAACCCAGCCGACGAGGAAAGCGCGCAGATGGCTGCTCGGCACACGTC 100
|||||
242 yasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGGAACCATCCGCTGGCTGTGTGATCAGGCCCGACGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGCTGCGCGCGGAGTCTACCTACCTGGCGCAGGTGGTCTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGlyProValAlaProSerValM 309
|||||
251 TCCCGGCGGCTGTTCCCGGATCGCTGAGCGGTGGCGGCTCCGCTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGCGATGGCGGAGGTTCGCAATCCCGCGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTGCGCGCGGACCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGGAGGAGGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-056-556-114

```
seq_documentation_block:
; Sequence 114, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
```

```
;
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Maki, David J.
;   REGISTRATION NUMBER: 31,392
;   REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 622-4900
;   TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 368 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; US-09-056-556-114
;
; alignment_scores:
;   Quality: 713.00      Length: 143
;   Ratio: 5.057        Gaps: 0
; Percent Similarity: 98.601 Percent Identity: 96.503
;
; alignment_block:
; US-09-462-480-3 x US-09-056-556-114
;
; Align seg 1/1 to: US-09-056-556-114 from: 1 to: 368
;
; 1 CTCGACGAGTGGTGTTCAGCGAGTGGCGGCGCGCGCGG 50
; |||||
; 226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
;
; 51 CAACCCAGCGAGCAGGAGCGCGGCGGAGTGGCGCTCTCGGCACCAAGTC 100
; |||||
; 242 YasnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
;
; 101 CGCTGTGCAACCATCCCTCGCTGGTGGATCATAGGCCCGCGCGGCGCG 150
; |||||
; 259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
;
; 151 GGCTGTGCTGCGCGGAGTGCCTACCTGCGCGAGTGGTGGTGGTACCCG 200
; |||||
; 276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
;
; 201 CAGCGCGCTGATGTCTCAGCTGATCGAAGCGGTTGCCCGCTCGGTGA 250
; |||||
; 292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerVal 309
;
; 251 TGCGCGCGCTGTTCGCGGATTCGCTGACGGGTGGCGCGCTCCGGTG 300
; |||||
; 309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaPro 325
;
; 301 GGTCCGGGAGCGATGGCCAGGTTCCCAATCCGCGCGCTCCACGCCC 350
; |||||
; 326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
;
; 351 GGGTCTGTGCGCGCGCGCGCGTCCGCGAGGAGCGTGAAGAGACGACG 400
; |||||
; 342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
;
; 401 AGGACGACTGGGAGGAGGAGGAGTGG 429
; |||||
; 359 luAspAspTrpAspGluGluAspTrp 368
;
; seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-09-010-928B-28
;
; seq_documentation_block:
; Sequence 28, Application US/09010928B
; Patent No. 5994099
;
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
```

```
;
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 745 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-010-928B-28
;
; alignment_scores:
;   Quality: 128.00      Length: 132
;   Ratio: 1.707        Gaps: 6
; Percent Similarity: 56.818 Percent Identity: 35.606
;
; alignment_block:
; US-09-462-480-3 x US-09-010-928B-28
;
; Align seg 1/1 to: US-09-010-928B-28 from: 1 to: 745
;
; 37 GGCACCGCGCGCGCGCGCGCGCGCGCGAGTGGCGCT 86
; |||||
; 251 GlyProGlyGlyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyPr 267
;
; 87 GCTCGGCGACCGAGTCCGCTGTCGAACCATCCGCTGGTGGTGATCAGGCC 136
; |||||
; 267 oGlyGlyTyrglyProGlyGlySerGlyPro.....GlyGlyAlaGlyP 282
;
; 137 CCAGCGCGCGCGCGCGCTGCTGCGCGGAGTGCCTACCTGGCGCAGGT 186
; |||||
; 282 roser...GlyAlaGlyLeuGlyGlyAlaGly.....ProGlyGlyAla 295
;
; 187 GGGTCTGTGACCCGCGCGCTGATGCTCAGCTGATCGAAGACCGCGT 236
; |||||
; 296 Gly..... 296
;
; 237 TGCCCCCTCGGTGATCGCGCGCTGTCGCGGATCGTCTGCGTACGGGTG 286
; |||||
; 297 .....LeuGlyGlyAlaGlyProGlyGlyAlaGlyThrSerG 309
;
; 287 GCGCGCTCCG.....GTGGTCCGGAGCGATGGCGCAGGTTCGCCAA 330
; |||||
; 309 lyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyGlnGlyAspAla 325
;
; 331 TCCGCGCGCTCCACGAGCC.....GGTCTGGTGGCGCGCGCGCGCTCG 376
; |||||
; 326 GlyProGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyValGlyAr 342
;
; 377 GCGAGGAGCGTGAAGAACGACGCGCGAGTGGGAGGAGAGAGA 422
; |||||
; 342 gGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyAlaArgGly 357
;
; seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-09-010-928B-2
```

[illegible]

36 aAlaAlaAlaGlyGlyProLeuArgTyrGlyArgAlaValArgAlaArgG 53
52 AACCC.....AGCCGAGGAGAGCCGCGCAGATGGCGCTGTCGCGCAC 95
53 lyProArgAspAlaArgGlyAla.....AlaProGlyArgGly 66
96 CAGTCCTGCTGCAACCATCGCTGGTGGTGGTATCATGCCGCCAGCGCGG 145
67 AlaArgSerAlaThrProGlyAlaGly.....AlaProAlaArgG1 80
146 G.....CGGGGCGCTGCTGCGCGCG 165
80 yHisArgAlaArgAlaProAlaProArgArgGlyProAlaAlaArgG 97
166 GAGTCGTACCTGCGCGCAGGTGGTTCGTTGACCGCGCAGCGCGTGATGTC 215
97 ly..... 97
216 TCAGCTGATCGAAAGCGGTGCCCCCTCGGTGATCGCGCGGCTGTG 265
98GlyArgGlyGlyProArgAlaGlyAlaLe 108
266 CCGGATCGTGGTGACGGTGGCGCC.....GCTCGGTGGGTCCG 306
108 uArgAlaGlyGlyArgGlyGlyAlaArgGlyProAlaGluGlyAla 124
307 GGAGCGATGGCCAGGTTCGCAATCCGCGGCTCCACCGCCCGG... 353
125 GlyAlaAlaGlyGlyValArgLeuProAlaAlaProProProGlyArg 140
354 TCTGTCGCGCGCGCGCTCGCGCAGGAGCGTGAAGAAGACGACGAGG 403
141 GlyGlyArgAlaAlaArgProAspProGlyLeuArgArgAlaGlyAl 157
404 ACGACTGGGAGGAGAGGAGACTGGTGGTACGTCCTCCGTAATGAC 446
157 aAspAlaGlyArgAspAlaArgArgProGluValArgArgAsp 171
seq_name: /cqn_6/ptodata/2/1aa/6B_COMB.pep:US-09-413-814-78

seq_documentation_block:
; Sequence 78, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bioecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 78
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-78

alignment_scores:
Quality: 125.00 Length: 138

Ratio: 1.786 Gaps: 10
Percent Similarity: 50.725 Percent Identity: 37.681
alignment_block:
US-09-462-480-3/rev x US-09-413-814-78 ..
Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882
424 CGTCTCTTCTGTCCTCCAGTCCTCTCTCTCTTTCACGCTCCTCGG... 377
347 ArgProThrArgProSerArgProArgArg.....ThrProSerAr 360
376CGAGCGGTGCGCGCGGACGACGACCGCGGTGTGGAGCGCGCGGA 331
360 gArgArgArgSerProAlaArg..... 367
330 TTGGGAACCTTGCCCATCGTCCCGGACCCACCGGAGCGGCGCCACCG 281
368ThrProGlyProArgSerPro.....ArgArgArgPro 378
280 TCACCGACGATCCGG.....CAACAGCGCGCGGCATCACC 246
379 SerProSerAlaArgSerProAlaIleTrpGluArgProProArgArgPr 395
245 GAGGGGCAACCGGCTTTTCATCAGCTGAGACATCAGCGCGTGCCTGGT 196
395 oArgAspArgArgProArgArg.LeuLeuArgProAlaArgArgLeuAla 411
195 CAAGAGCCACCT.....GCGCCAGTAGGACTCCGCGCGCAGCAGGC 152
412 ArgArg.AspproGlyAspValProAspProArgAlaAlaArgArgArg. 427
151 CCGCGCGCGCTGGCGCTGATCCACCGCAGCGGATGTTTCGACAGC 102
428 ..AlaProAlaAlaGlyAlaLeuProGlyAlaAspGlyArgGlyAlaArg 443
101 GGACTGTGTCGAGCAGCGCCCATCTGCGCGGCTTCTCTCGTGGGTGTT 52
444 Gly.....ProGlyArgArgArgAlaProArgArgGlyAlaAlaAr 458
51 GCGCGCGCGCG 42
458 gValProPro 461

seq_name: /cqn_6/ptodata/2/1aa/6B_COMB.pep:US-09-232-468A-8

seq_documentation_block:
; Sequence 8, Application US/09232468A
; Patent No. 6207165
; GENERAL INFORMATION:
; APPLICANT: AUDONNET et al.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY PATHOLOGIES
; FILE REFERENCE: 454313-2230
; CURRENT APPLICATION NUMBER: US/09/232,468A
; CURRENT FILING DATE: 1999-01-05
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Aujeszky's Disease Virus (NIA3 Strain)
US-09-232-468A-8

alignment_scores:
Quality: 121.00 Length: 166
Ratio: 1.424 Gaps: 7
Percent Similarity: 51.205 Percent Identity: 30.723

alignment_block:
US-09-462-480-3/rev x US-09-232-468A-8 ..

;; TITLE OF INVENTION: Modulators of Actin
;; NUMBER OF SEQUENCES: 14
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
;; STREET: 75 DENISE DRIVE
;; CITY: HILLSBOROUGH
;; STATE: CALIFORNIA
;; COUNTRY: USA
;; ZIP: 94010
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/323,735
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/080,897
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: OSMAN, RICHARD A.
;; REGISTRATION NUMBER: 36,627
;; REFERENCE/DOCKET NUMBER: UW97-001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (650) 343-4341
;; TELEFAX: (650) 343-4342
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1248 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-09-323-735-2

alignment_scores:
Quality: 120.50 Length: 149
Ratio: 1.826 Gaps: 8
Percent Similarity: 44.295 Percent Identity: 33.557

alignment_block:

US-09-462-480-3/rev x US-09-323-735-2 ..
Align seg 1/1 to: US-09-323-735-2 from: 1 to: 1248
448 TTGTCATTACGGAGCTCACCAGTCGTCTTCTGTCCTCCAGTCGTCTCG 399
||||| :||| ||| |||||
556 LeuSerAlaAlaAlaIleThrValProProSerValProSerArg..... 570
398 TCGTCTTCTCAGGCTCTCGCGGAGCGGTGCGCGCGGACGACACCGG 349
||| :||| ||| |||||
571AlaProValProProAlaProProLeuProG 581
348 GGTGTGGAGCGCGGATGCGAACCTGCGCCATCGCTCCCGACCCA 299
|| :||| ||| :||| |||
581 Ly.....AspSerGlyThrIleIleProPro 590
298 CCG.....GACGGCGCCACCGTCACGACGATCCGGCAACAGCC 258
||| :||| ||| |||
591 ProAlaProGlyAspSerThrProProProProProProProPr 607
257 GCCGCATCACCGAG.....GGGGCAACGGGTTTTCGATCAGCTAGACA 212
||||| ||| ||| :||| :||| :|||
607 oProProProProLeuProGlyThrAlaIleSerProProProPro 624
211 TCACGGCGGTGCGGTCAACGACCACTCGGCCA..... 177
:||| ||| :||| |||
624 euSerGlyAspAlaThrIleProProProProProProGluGlyVal 640
176 GTATCGACTCGCGCGGACGAGGCC...GCGCCGCGCTGGGCGCTGA 130
||| ||| ||| ||| :||| :|||

641 GlyIleProSerProSerSerLeuProGlyGlyThrAlaIleProProPr 657
129 TCCACCGCCAGCGATGTTCCGACGCGGACTGTGTCGCGGAGGCCCA 80
||||| ||| ||| :||| :||| |||
657 oProProLeuProGly.....SerAlaArgIleProProProPro. 670
79 TCTGCGCGGCTTCTCGTCGCTGGTGGTGGCGCGCGGTCGCCGCC 33
:||||| :||| :||| :||| :||| :|||
671 ..ProProLeuProGlySerAlaGlyIleProProProProProPro 685
seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-899-595-3
seq_documentation_block:
; Sequence 3, Application US/08899595
; Patent No. 6111072
; GENERAL INFORMATION:
; APPLICANT: Narumiya, Shuh
; APPLICANT: Takahashi, No. 6111072uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
; TITLE OF INVENTION: ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,595
; FILING DATE: 24-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-242701
; FILING DATE: 26-AUG-1996
; PRIOR APPLICATION DATA: JP 9-90170
; APPLICATION NUMBER: JP 9-90170
; FILING DATE: 25-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stephen A. Bent
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 049441/0112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-899-595-3

alignment_scores:

Quality: 120.00 Length: 156
Ratio: 1.714 Gaps: 8
Percent Similarity: 44.872 Percent Identity: 32.051

alignment_block:

US-09-462-480-3/rev x US-08-899-595-3 ..
Align seg 1/1 to: US-08-899-595-3 from: 1 to: 1315
448 TTGTCATTACGGAGCTCACCAGTCGTCTTCTGTCCTCCAGTCGTCTCG 399
||||| :||| ||| |||||
608 LeuSerAlaAlaAlaIleThrValProProSerValProSerArg..... 622

```
398 TCGTCTTCTTACGGCTCT..... 380
623 .....AlaProValProAlaProProLeuProGlyAspS 635
379 ..GCGGAGCGGTGCGCGGACACCGCGGTGTGTGGAGCGCGG 332
635 erGlyThrIleileProProAlaProGly.....AspSer 648
331 ATTGGAAACCTGGCCCTCGCTCCGACCCACCGGAGCGGCCACCC 282
649 ThrThrProProProProProProProProProProProPro 665
281 GTCACCGACGATCCGCGCAACCGCGCATCCAGAGGCGCAACCGG 232
665 oGlyGlyValCysIleSerSerProProSerLeuPro..GlyGlyThrAl 681
231 CTTTTCATCAGCTGAGATCAGCGGTGCGGTCAACACCCACCTG 182
681 aileSerProProProProLeuSerGlyAspAlaThrIleProPro 698
181 CGCCA.....GGTAGCGACTCGCGCGCAGCAGGCC... 150
698 roProLeuProGluGlyValGlyIleProSerProSerSerLeuProGly 714
149 GCGCGCGCGTGGCGCTGATCCACCGACCGAGCGATGTTTCGACAGCGG 100
715 GlyThrAlaIleProProProProProProGly.....SerAl 728
99 ACTGGTCGCGAGCAGCGCATCTGCGCGGTCTCTCGTGGGTGGTGGC 50
728 aArgIleProProProPro...ProProLeuProGlySerAlaGlyIleP 744
49 CGCGCGCGTGGCGGCC 33
744 roProProProPro 749

seq_name: /cgn2_5/pdata/2/1aa/6B_COMB.pep:US-08-483-533-41

seq_documentation_block:
; Sequence 41, Application US/08483533
; Patent No: 6172047
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; APPLICANT: Chou, Joany
; TITLE OF INVENTION: Method for Treating Tumorigenic
; TITLE OF INVENTION: Diseases
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Bdrun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,533
; FILING DATE: 07-MAR-95
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/419,853
; FILING DATE: 11-APR-95
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/861,233
; FILING DATE: 31-MAR-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, James P.
; REGISTRATION NUMBER: 28,491
```

```
; REFERENCE/DOCKET NUMBER: 28097/32742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
```

```
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-483-533-41
```

```
alignment_scores:
  Quality: 118.50      Length: 222
  Ratio: 1.411        Gaps: 9
  Percent Similarity: 37.838  Percent Identity: 26.126
```

```
alignment_block:
```

```
US-09-462-480-3/rev x US-08-483-533-41 ..
```

```
Align seg 1/1 to: US-08-483-533-41 from: 1 to: 355
```

```
440 ACGGAGCT.....CACCAGTCGTCTCTTCGTCCTCCAGTCGTCTC 400
19 ThrGlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSe 35
399 GTCTGCTTCTTACGCTCTCGCGGAGCGGTGCGCGGACGACGCGG 350
35 rGluProAlaValArgSerAlaProAlaAlaProProProProA 52
349 GGCTGTGGAGCCGCGGATTGC..... 327
52 laGlyGlyProProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis 68
326 .....GAACCTGGCCCATCGC 310
69 ValProGluSerAlaSerAspAspAspAspAspTrpProAspSe 85
309 TCCC.....GGACCCACGAGCGCGG.... 288
85 rProProGluSerAlaProGluAlaArgProThrAlaAlaProA 102
287 ..GCACCGTCCACGACGATCGGCA.....ACA 261
102 rgProGlyProHisArgProAlaTrpAlaArgGlyAlaGlyLeuThr 118
260 GCGCGCGCATCACGAGGGGCAACCGCTTTTCGATCAGCTGAGACAT 211
119 .ProProThrProProArgAlaProSerAlaPheArgAlaSerProS 135
210 CAGCGCGTGGGTACAGCACCCACCTGCGCGGACGAGTACGACTCCGCG 161
135 erAlaCysAlaSerProArgSerThrTrpArgAlaCysAlaCysAspAla 151
160 GCACGAGC..... 152
152 ArgAlaGlyArgGlyArgSerProProArgProProArgProProAr 168
151 .....CGCGCGCGCTGGGCGCTGATCCACCAGCGCGG 115
168 gProProArgProProArgProProArgGlyCysAlaSerArgProThrS 185
114 ATGGTTCGACAGCGGACTGTGCGCGGACGAGCGCCCATCTGCG.... 74
185 erGlyCysAlaThrTrpTrpSerGlyProArgProProAlaTrpArgAla 201
74 ..... 74
202 AlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySerGlyAlaGl 218
```

```

19 ThrGlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSe 35
399 GTCGGTCTTCTTACGCTCTCTGCGGACGAGCGGTGCGGCGGACGACACCG 350
    :::: |||||::: ::::: |||:
35 rcluProAlaValArgSerAlaProAlaAlaAlaProProProProA 52
349 GCCTGGTGAGCGCGCGGATTGC..... 327
    :: |||||::: |||
52 laGlyGlyProProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis 68
326 ..... GAACCTCGGCCCATCGC 310
69 ValProGluSerAlaSerAspAspAspAspTrpProAspSe 85
309 TCCC..... GGACCCACCGGAGCGCG... 288
    :||| |||||::: |||||
85 rProProGluSerAlaProGluAlaArgProThrAlaAlaAlaProA 102
287 ..CCACCCGTACCGAGGATCCGGCA.....ACA 261
    ||||| |||||
102 rgProProGlyProHisArgProAlaTrpAlaArgGlyAlaGlyLeuThr 118
260 GCCCGCGGCATCACCGAGGGGCAACCGCTTTTCGATCAGCTGAGACAT 211
    |||||::: |||||::: |||||::: |||||:::
119 .ProProTrpProProArgAlaProSerAlaPheArgArgAlaSerPro 135
210 CAGCGCGGTGCGGGTCAACGACCCACCTGGCGCCAGGTAGCGACTCCGCGC 161
    |||||::: ::: |||
135 eAlaCysAlaSerProArgSerThrTrpArgAlaCysAlaCysAspAla 151
160 GCAGCAGGC..... |||||
152 ArgAlaGlyArgGlyArgArgSerProProArgProProArgProProAr 168
151 ..... CCGCGCCGCGCTGGGCGCTCATCCACGACGACGCGG 115
168 gProProArgProProArgProProArgGlyCysAlaSerArgProThrs 185
114 ATGGTTCGACAGCGGACTGGTGGCGAGCAGCGCCCATCTGCG..... 74
    :||| ::::: ||| |||
185 eRGlyCysAlaThrTrpTrpSerGlyProArgProProAlaTrpArgAla 201
74 ..... 74
202 AlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySerGlyAlaGl 218
73 ....CGGCTCTCTGCTGGCTGGTGGTGG.....CGCGCGCGCGG 41
    ||| ||||| ||| ::: ||| ||||
218 yTrpArgProProArgArgSerSerGlyArgAlaTrpGlyProArgProV 235
40 TGCGCGCCACCTGG 27
    ||||| |||
235 alProGlyProTrp 239

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US91-06532-3
seq_documentation_block:
; Sequence 3, Application PC/TUS9106532
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: Recombinant Herpes Simplex Viruses
; NUMBER OF INVENTIONS: Vaccines and Methods
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza Suite 2100
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/06532
; FILING DATE: 19910910
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gruber, Lewis S.
; REGISTRATION NUMBER: 30,060
; REFERENCE/DOCKET NUMBER: 27373/8235
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/346-5750
; TELEFAX: 312/984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US91-06532-3

alignment_scores:
  Quality: 118.50      Length: 222
  Ratio: 1.411        Gaps: 9
  Percent Similarity: 37.838   Percent Identity: 26.126

alignment_block:
US-09-462-480-3/rev x PCT-US91-06532-3  ..
Align seg 1/1 to: PCT-US91-06532-3 from: 1 to: 355

440 ACCEGAGCT.....CACAGTCGCTCTCGTCCAGTCGCTCCTC 400
|||||
19 ThrGlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSe 35
399 GTCGTCCTTTCACGCTCTCGGAGCGGTGCGCGCCGACGACCG 350
|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
35 rGluProAlaValArgSerAlaProAlaAlaProProProA 52
349 GCCTGGTGGAGCCCGCGATTGC..... 327
: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
52 IaGlyGlyProProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis 68
326 .....GAACCTGGCCCATCGC 310
|||: :|||: :|||: :|||: :|||: :|||: :|||:
69 ValProGluSerAlaSerAspAspAspAspAspTrpProAspSe 85
309 TCCC.....GGACCCACCGAGCGG..... 288
: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85 rProProGluSerAlaProGluAlaArgProThrAlaAlaProA 102
287 ..CCACCGTCACCGACATCCGCA.....ACA 261
||||| |||||
102 rgProProGlyProHisArgProAlaTrpAlaArgGlyAlaGlyLeuThr 118
260 GCGCGCGGCATCACGAGGCGGACCGCTTTTCGATCAGTCAGACAT 211
|||||: :|||: :|||: :|||: :|||: :|||: :|||:
119 .ProProThrProProArgAlaProSerAlaPheArgArgAlaSerProS 135
210 CAGCGGGTGGGTCAACAGCCACCTGCGCCAGGTAGCGACTCCGCGC 161
|||||: :|||: :|||: :|||: :|||: :|||: :|||:
135 eAlaCysAlaSerProArgSerThrTrpArgAlaCysAlaCysAspAla 151
160 GCACGAGC..... 152
|||||
152 ArgAlaGlyArgGlyArgSerProProArgProProArgProProAr 168
151 .....CCGCGCGCGCGTGGGCGCTGATCCACCGACGCGG 115
|||||
168 gProProArgProProArgProProArgGlyCysAlaSerArgProThrS 185
```

```

114 ATGGTTGACAGCGGACTGGTCCGAGCAGGCCATCTGCG..... 74
: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
185 eGlyCysAlaThrTrpSerGlyProArgProAlaTrpArgAla 201
74 ..... 74
202 AlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySerGlyAlaGl 218
73 .....CGGCTTCTCGCTCGCTGGGTG.....CGCCGCCGG 41
218 yTrpArgArgProArgArgSerGlyArgAlaTrpGlyProArgProV 235
40 TGCCGCCACCTGG 27
235 alProGlyProTrp 239
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-09-010-928B-4
seq_documentation_block:
; Sequence 4, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010, 928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-010-928B-4
```

```

alignment_scores:
  Quality: 118.00      Length: 119
  Ratio: 1.662        Gaps: 6
  Percent Similarity: 59.664   Percent Identity: 35.294

alignment_block:
US-09-462-480-3 x US-09-010-928B-4  ...
Align seg 1/1 to: US-09-010-928B-4 from: 1 to: 907

37 GGCACCGCGCGCGCAACCCAGCGAGGAGCGCGCAGATGGGCGT 86
||| |||||: :|||: :|||: :|||: :|||: :|||:
186 GlyProGlyGlyAlaGlyProGlyGlyAlaGlyProGly 202
87 GCTCGGACCAAGTCGCTGTGCAACCATCCGCTGGTGGTATCAGGCC 136
|||: :|||: :|||: :|||: :|||: :|||: :|||:
|||||: :|||: :|||: :|||: :|||: :|||: :|||:
```

202 oGlyGlyAlaGlyProGlyGlyAlaGlyPro.....GlyGlyAlaGlyG 217
137 CCAGCGCGGCGCGGCGCTGCTGCGCGGAGTCGCTACCTGGCGCAGGT 186
: : : : : ||||| ||| : : : |||||
217 LyAlaGlyGlyAlaGly.....GlyAlaGlyGlySerGlyGlyAlaGly 231
187 GGGTCGTGACCCGCGCGCTGATGTCACGTG..... 222
||||| ||| : : : : : |||||
232 GlySer...GlyGlyThrThrIleIleGluAspLeuAspIleThrIleAs 247
223ATCGAAAGCGGTTGCCCCCTCG...GTGATGCCGCGGCTGTTG 265
: : : : : ||||| ||| : : : : : |||||
247 pGlyAlaAspGlyProIleThrIleSerGluGluLeuThrIleSerGlyA 264
266 CCGATATCGGTGACGGTGGCGCGCTCCG.....GTGGTCCGGGA 309
||| : : : ||| ||||| ||||| ||||| ||||| |||||
264 IaGlyGlySerGlyProGlyGlyAlaGlyProGlyGlyValGlyProGly 280
310 CGGATGGCGCGGTTCCGAATCCGGCGGCTCCACCGCGGCTCTGCT 359
: : : ||| : : : : : ||||| ||||| ||||| |||||
281 GlySerGlyProGlyGlyValGlyProGlyGlySerGlyProGlyGlyVa 297
360 CGCACC 366
: : : |||||
297 IGlyPro 299

THIS PAGE BLANK (USPTO)

151 GGCTGTGTCGGCGGAGTCGCTACCTCGCGCAGGTGGGTGCTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
|||||
201 CACCCGCTGATGCTCAGCTGATCGAAAGCCGGTTCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGCGGGCGGCTGTTCGGGATCGTGGGTGACGGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGGATGGCGCAGGTTCGAATCCGGCGCTCCACAGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTGTCGGCGGCAACGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
|||||
401 AGGACGACTGGGAGCAGGAGCAGCTGG 429
359 luAspAspTIPAspGluGluAspAspTIP 368

seq_name: /cgn2_5/ptodata/2/paa/PCTUS_COMB.pep.PCT-US99-03268-114

seq_documentation_block:
; Sequence 114, Application PC/TUS9903268
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US99/03268
; FILING DATE: 17-FEB-1999
; CLASSIFICATION: 406
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/025,197.
; FILING DATE: 18-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9532-0013-228
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
PCT-US99-03268-114

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-3 x PCT-US99-03268-114

Align seg 1/1 to: PCT-US99-03268-114 from: 1 to: 368
1 CTGCAGCAGGTGACGTCGTTGTTACGCCAGGTGGCGGCACCGCGGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
|||||
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGGCTGCTCGGCACACGTC 100
|||||
242 yasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCGAACCATCCGCTGGGTGGTGGATCAGGCCCGCAGCGCGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGTGTCGGCGGAGTCTGCTACCTGGCGCAGGTGGGTGCTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCGGTTCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGCGGGCGGCTGTTCGGGATCGTGGGTGACGGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGGATGGCGCAGGTTCGAATCCGGCGCTCCACAGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTGTCGGCGCAGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGAGCAGGAGCAGCTGG 429
359 luAspAspTIPAspGluGluAspAspTIP 368

seq_name: /cgn2_5/ptodata/2/paa/US086_COMB.pep.US-08-658-800-109

seq_documentation_block:
; Sequence 109, Application US/08658800
; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS

; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,800
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C3
; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-658-800-109

```

```

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

```

```

alignment_block:
US-09-462-480-3 x US-08-658-800-109

```

```

Align seg 1/1 to: US-08-658-800-109 from: 1 to: 368

```

```

1 CTCAGCAGGTGACGCTGTTGTCAGCCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
51 CAACCCAGCGCAGGAGGAGCGCGCAGATGGGCTGCTCGGCACCAAGTC 100
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCCGCTGGTGGTGGATCAGGCCCGCGCGCGCG 150
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCTGCTGCGCGCGGAGTGCTACCTACCTGCGCAGGTGGTGGTGG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
201 CACGCCGCTGATGTCTCAGCTGATCGAAGACCGGTTGCCCGCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTGCGGAGTCGCTGGTGGTGGTGGTGGTGGTGGTGG 300
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGGAGCGATGGCCAGGTTCCCAATCCGCGCGGCTCCACCAAGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlySerThrArgPr 342
351 GGCTGCTGTCGCGCGCACCGCTCGCGCAGGAGCGGTGAAGAACAGCAG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
401 AGGACGACTGGGACGAGAGGAGGACTGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368

```

```

seq_name: /cgn2_6/ptodata/2/paa/US086_COMB.pap:US-08-659-683-114

```

```

seq_documentation_block:
; Sequence 114, Application US/08659683
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington

```

```

; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,683
; FILING DATE: 05-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-659-683-114

```

```

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

```

```

alignment_block:
US-09-462-480-3 x US-08-659-683-114

```

```

Align seg 1/1 to: US-08-659-683-114 from: 1 to: 368

```

```

1 CTCAGCAGGTGACGCTGTTGTCAGCCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
51 CAACCCAGCGCAGGAGGAGCGCGCAGATGGGCTGCTCGGCACCAAGTC 100
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCCGCTGGTGGTGGATCAGGCCCGCGCGCGCG 150
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCTGCTGCGCGCGGAGTGCTACCTACCTGCGCAGGTGGTGGTGG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
201 CACGCCGCTGATGTCTCAGCTGATCGAAGACCGGTTGCCCGCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTGCGGAGTCGCTGGTGGTGGTGGTGGTGGTGGTGG 300
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGGAGCGATGGCCAGGTTCCCAATCCGCGCGGCTCCACCAAGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlySerThrArgPr 342
351 GGCTGCTGTCGCGCGCACCGCTCGCGCAGGAGCGGTGAAGAACAGCAG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
401 AGGACGACTGGGACGAGAGGAGGACTGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368

```

seq_name: /cgn2_6/ptodata/2/paa/US086_COMB.pep:US-08-680-573-109

seq_documentation_block:
; Sequence 109, Application US/08680573
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,573
; FILING DATE: 12-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-680-573-109

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x US-08-680-573-109 ..
Align seg 1/1 to: US-08-680-573-109 from: 1 to: 368

1 CTCGACGAGGTGAGTCTGTTTTCAGGCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242

51 CAACCCAGCCGACGAGGAGCGCGAGATGGCCTGCTCGGCACCAATC 100
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTGACCATCCCTCGTGGTGGATCAGGCCCGGCGGCGG 150
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275

151 GGCTGTGTCGCGCGGAGTCTGCTACCTGGCGGCGAGTGGTGGTGG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292

201 CAGCCGCTGATGCTCAGCTGATCGAAAGCCGGTTGCCCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309

251 TGC CGCGCGGTGTTGCCGATCGTCGGTGACGGTGCGCGGCTCCGGTG 300
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGGAGCATGGCGGAGGTTCGCAATCCGGCGGCTCCACGAGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
351 GGGTCTGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
401 AGGACGACTGGGACGAGGAGGAGGAGGACTGG 429
359 luAspAspTrpAspGluGluAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US086_COMB.pep:US-08-680-574-114

seq_documentation_block:
; Sequence 114, Application US/08680574
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 133
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,574
; FILING DATE: 12-JUL-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-680-574-114

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x US-08-680-574-114 ..
Align seg 1/1 to: US-08-680-574-114 from: 1 to: 368

1 CTCGACGAGGTGAGTCTGTTTTCAGGCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242

```
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGGCTGCTCGGCACCAAGTC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCTGAACCATCCGCTGCTGGTGGATCAGGCCCGCGGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGTCTGCGCGGAGTCTACCTGCGCAGGTGGTGGTTCGTTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCGTGATGCTCAGCTGATCGAAGCCGCTGCGCCCTCGGTGTA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309
|||||
251 TGCCGGGGCTGTCGCGGATCGCTGAGCGGTGGCGCGCTCGCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCTGCTGCGCGGAGTTCGCAATCCGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGTCTGCTGCGCGGACCGCTCGCGCAGAGCGTGAAGAACACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGAGGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US087_COMB.pap:US-08-729-622-109

```
seq_documentation_block:
; Sequence 109, Application US/08729622
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas H.
; APPLICANT: Twardzik, David R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 132
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/729,622
; FILING DATE: 11-OCT-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-729-622-109

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x US-08-729-622-109
Align seg 1/1 to: US-08-729-622-109 from: 1 to: 368

1 CTGACGACGAGTGACGCTGTTTTCAGCCAGGTGGCGGCACCGCGGGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
|||||
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGGCTGCTCGGCACCAAGTC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCTGAACCATCCGCTGCTGGTGGATCAGGCCCGCGGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGTCTGCGCGGAGTCTGCTGAGCGGTGGCGAGGTGGTTCGTTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCGTGATGCTCAGCTGATCGAAGCCGCTGCGCCCTCGGTGTA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309
|||||
251 TGCCGGGGCTGTCGCGGATCGCTGAGCGGTGGCGCGCTCGCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCTGCTGCGCGGAGTTCGCAATCCGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGTCTGCTGCGCGGACCGCTCGCGCAGAGCGTGAAGAACACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGAGGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||

seq_name: /cgn2_6/ptodata/2/paa/US087_COMB.pap:US-08-730-510-114

seq_documentation_block:
; Sequence 114, Application US/08730510
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 137
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
```

```
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/730,510
; FILING DATE: 27-AUG-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-730-510-114

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
  US-09-462-480-3 x US-08-730-510-114

  Align seg 1/1 to: US-08-730-510-114 from: 1 to: 368

1 CTGAGCAGGTGACGTCGTTGTTACGCCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
51 CAACCCAGCCGACGAGGAGCGCGGCACATGGCTGCTCGGCACCATGC 100
242 yasnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCCGCTGGCTGGTGCATCAGGCCCGCGCGCGCG 150
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGGAGTCGCTACCTGGCGCAGGTGGGTGCTGACCCG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
201 CACGCCGTGATGCTCAGCTGATCGAAGACCGGTTCCTCCCTCGGTGA 250
292 gthrProLeuMetSerGlnLeuIleGluProValAlaProSerValM 309
251 TCGCGCGCGCTGTTCCCGATGCTCGGTGACGGGTGGCGCGCTCCGGTG 300
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGAGCGATGGGCGCAGGGTTCGCAATCCCGCGGCTCCACAGCCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgpr 342
351 GGGTCTGTCGCGCGCGACCGCTCGCGCAGGAGCGGTGCAAGACGACG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspaspG 359
401 AGGACGACTGGGACGAAGAGGACGACTGG 429
359 luAspAspTrpAspGluAspAspTrp 368
```

```
seq_name: /cgn2_6/ptodata/2/paa/us089_COMB.pep.us-08-942-341-109

seq_documentation_block:
; Sequence 109, Application US/08942341
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,341
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-942-341-109

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
  US-09-462-480-3 x US-08-942-341-109

  Align seg 1/1 to: US-08-942-341-109 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGTTGTTACGCCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
51 CAACCCAGCCGACGAGGAGCGCGGCACATGGCTGCTCGGCACCATGC 100
242 yasnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCCGCTGGCTGGTGCATCAGGCCCGCGCGCGCG 150
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGGAGTCGCTACCTGGCGCAGGTGGGTGCTGACCCG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
201 CACGCCGTGATGCTCAGCTGATCGAAGACCGGTTCCTCCCTCGGTGA 250
292 gthrProLeuMetSerGlnLeuIleGluProValAlaProSerValM 309
251 TCGCGCGCGCTGTTCCCGATGCTCGGTGACGGGTGGCGCGCTCCGGTG 300
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGAGCGATGGGCGCAGGGTTCGCAATCCCGCGGCTCCACAGCCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgpr 342
351 GGGTCTGTCGCGCGCGACCGCTCGCGCAGGAGCGGTGCAAGACGACG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspaspG 359
401 AGGACGACTGGGACGAAGAGGACGACTGG 429
359 luAspAspTrpAspGluAspAspTrp 368
```

```
201 CACGCCGCTGATGCTCAGCTGATCGAAGAACCGGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TCCCGCGCGCTGTCGCCGATCGCGTGACGGGTGGCGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaMetGlySerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGGAGTGGCGAGGCTCGCAATCCGGCGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlySerThrArgPr 342
|||||
351 GGGTCTGTCGCGCGCGCACCGCTCGCGAGAGCGTGAAGAACGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359
|||||
401 AGGAGCTGGGACGACGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US089_COMB.pep:US-08-942-578-114

```
seq_documentation_block:
; Sequence 114, Application US/08942578
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,578
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-942-578-114
```

alignment_scores: Quality: 713.00 Length: 143
 Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

```
alignment_block:
US-09-462-480-3 x US-08-942-578-114
Align seg 1/1 to: US-08-942-578-114 from: 1 to: 368

1 CTGACAGAGTGACGTGCTGTTTCAGCACAGTGGCGGCGCACCGCGCGGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
|||||
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGGCTGCTCGGCGACCACTC 100
|||||
242 yAspProAlaAspGluGluAlaAlaGlnMetGlyLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGCTGCTGATCAGGCCAGGCCCGCGGGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGTGCGCGCGGAGTCTGCTACCTGCGCGCAGTGGGTGCTTGACC 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAGAACCGGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TCCCGCGCGCTGTTGCCGATCGCTGCGTGACGGGTGGCGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGGAGTGGCGCGAGGCTCGCAATCCGGCGCGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTCGCGCGCGCACCGCTCGCGAGAGCGTGAAGAACGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359
|||||
401 AGGAGCTGGGACGACGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-024-753-109

```
seq_documentation_block:
; Sequence 109, Application US/09024753
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 236
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/024,753
; FILING DATE: 18-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
```

TUBERCULOSIS

```
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-024-753-109

alignment_scores:
    Quality: 713.00      Length: 143
    Ratio: 5.057         Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x US-09-024-753-109 ..

Align seg 1/1 to: US-09-024-753-109 from: 1 to: 368

1 CTGCAGCAGGTGACGTGCTGTTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242

51 CAACCCAGCGCAGGAGGAGCGCGCAGATGGCCCTGCTCGGCACGATC 100
|||||
242 YasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGCCCGCAGCGCGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCCTGTGCTGGCGCGAGTCGCTACTTGCAGCAGGTGGCGGCACCGCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292

201 CAGCGCGCTGATGCTCAGCTGATCGAAAGCCGGTTGCCCTCCGCTGTA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGlySerProValAlaProSerValM 309

251 TGCCGCGCGGCTGTTGCCGGATCGTGGTGACGGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325

301 GGTCCGGAGCGATGGCGCAGGGTTTCGCAATCCGCGCGCTCCACCGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342

351 GGGTGTGTCGCGCGCGCAGCGCTCGCGCAGGAGCGGTGAAGAAGCAGCG 400
|||||
342 oglyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359

401 AGGACGAGCTGGGACGAGGAGGAGCTG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB pep: US-09-025-197-114

seq_documentation_block:
; Sequence 114, Application US/09025197
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
```


351 GGGTCTGGTCCGCGGACCGCTCGGCGAGGAGCGTGAAGAACGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGAGGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-072-596-109

seq_documentation_block:
; Sequence 109, Application US/09072596
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Veddzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-109

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-3 x US-09-072-596-109 ..
Align seg 1/1 to: US-09-072-596-109 from: 1 to: 368
1 CTGACGAGGTGACGTGCTGTTTCAGCCAGTGGCGGACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
|||||
51 CAACCCAGCCGACGAGGAGCGGCGGACATGGGCTGCTCGGACCGACTC 100
|||||

242 yAanProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACCATCCGCTGGTGGATCAGGCCCGCCAGCGGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGTGTCGCGCGGAGTCTACCTGCGCGCAGGTGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CAGCCCGCTGATGCTCAGCTGATCGAAAGCGGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309
|||||
251 TGC CGCGGCGCTGTTCCGCGATCGTCGTCGACGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGCGGAGCGATGGCCAGGGTTCGCAATCCGCGCGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGCTCTGTCGCGCGGCGCGCTCGCGCAGGCGTGAAGAACGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGAGGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||
seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-072-967-114
seq_documentation_block:
; Sequence 114, Application US/09072967
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Veddzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids

```
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-114

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
  US-09-462-480-3 x US-09-072-967-114

Align seg 1/1 to: US-09-072-967-114 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGTTGTCAGCCAGGTGGCGGCACCGCGCGG 50
  |||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAAGCGCGCAGATGGCTGCTCGGCACCGTC 100
  |||||
242 YasnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCCGCTGGCTGGATCATGAGCCCGCGCGCGCG 150
  |||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCTGCTGCGCGCGGAGTCGCTACCTGGCGCAGTGGTGGTGACCG 200
  |||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
201 CACGCCGCTGATGTCAGCTGATCGAAGCGCGGTGCCCTCGGTGA 250
  |||||
292 gThrProLeuMetSerGlnLeuLeuGluProValAlaProSerValm 309
251 TGCGCGCGCGTGTGCGCGGATCGTCGGGACGGTGCGCGCGCGTGG 300
  |||||
309 etProAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCGGGAGCGATGGCGGCGGTTCGCAATCCGCGCGCTCCACGCGC 350
  |||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlySerThrArgPr 342
351 GGGTGTGTCGCGCGCGCGCTCGCGCAGGAGCGGTGAAGAGCAGCG 400
  |||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGAGCGACTGGGACGAGAGCAGCGACTGG 429
359 luAspAspTrpAspGluGluAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pap.us-09-724-685-114

seq_documentation_block:
; Sequence 114, Application US/09724685
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir
; Dillion, Davin C.
; Campos-Neto, Antonio
; TITLE OF INVENTION: Compounds and Methods for
; Immunotherapy and Diagnosis of Tuberculosis
; NUMBER OF SEQUENCES: 155
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

[illegible]

THIS PAGE BLANK (USPTO)

Sequence	Strd Orig	zScore	EScore	Len	Documentation
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-080-170B-13	198.00	228.73	5.1e-03	
/cgn2_5/ptodata/2/paa/US09_NEW_COMB.pep	US-09-994-404-252	129.00	133.39	1.03	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-155-881-20236	127.50	150.53	1.34	
/cgn2_5/ptodata/2/paa/US60_NEW_COMB.pep	US-60-382-898-278	125.50	140.64	1.76	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-155-881-25973	125.00	146.21	1.92	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-080-170B-452	123.00	146.54	2.57	
/cgn2_5/ptodata/2/paa/PCT_NEW_COMB.pep	PCT-US02-09921-1022	121.50	140.91	3.17	
/cgn2_5/ptodata/2/paa/US60_NEW_COMB.pep	US-60-360-039-3962	121.50	136.92	3.14	
/cgn2_5/ptodata/2/paa/US09_NEW_COMB.pep	US-09-784-984-8	121.00	139.87	3.40	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-155-881-11183	121.00	136.02	3.37	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-155-881-10047	120.50	140.81	3.67	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-153-881-10704	120.50	140.28	3.66	
/cgn2_5/ptodata/2/paa/US09_NEW_COMB.pep	US-09-573-635B-274	119.50	140.17	4.24	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-153-881-7219	119.00	139.73	4.56	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-153-881-26034	118.50	138.18	4.89	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-105-239-6163A	117.00	135.35	6.05	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-160-568-2	117.00	129.40	5.97	
/cgn2_5/ptodata/2/paa/US09_NEW_COMB.pep	US-09-540-209B-10118	116.50	139.26	6.55	
/cgn2_5/ptodata/2/paa/US09_NEW_COMB.pep	US-09-667-170A-187	116.50	131.85	6.46	
/cgn2_5/ptodata/2/paa/US60_NEW_COMB.pep	US-60-360-039-19227	116.50	130.28	6.44	
/cgn2_5/ptodata/2/paa/PCT_NEW_COMB.pep	PCT-US02-16676-9	116.50	126.16	6.38	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-155-881-10185	116.00	132.99	6.97	
/cgn2_5/ptodata/2/paa/PCT_NEW_COMB.pep	PCT-US02-13874-3	115.50	126.72	7.40	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-155-881-7239	115.00	134.77	8.11	
/cgn2_5/ptodata/2/paa/US60_NEW_COMB.pep	US-60-360-039-2275	115.00	129.61	8.01	
/cgn2_5/ptodata/2/paa/US60_NEW_COMB.pep	US-60-360-039-20267	115.00	128.51	7.99	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-155-881-37314	115.00	128.33	7.99	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-155-881-97168	114.50	135.88	8.75	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-137-337-699	114.50	133.55	8.70	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-168-659-13	114.50	124.10	8.52	
/cgn2_5/ptodata/2/paa/US09_NEW_COMB.pep	US-09-819-104A-2	114.50	118.58	8.41	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-138-145-850	113.50	138.11	10.19	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-155-881-11387	113.00	135.39	10.90	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-155-881-9390	113.00	131.71	10.80	
/cgn2_5/ptodata/2/paa/US60_NEW_COMB.pep	US-60-360-039-4031	113.00	122.29	10.57	
/cgn2_5/ptodata/2/paa/PCT_NEW_COMB.pep	PCT-US02-09944-546	112.50	125.33	11.46	
/cgn2_5/ptodata/2/paa/US60_NEW_COMB.pep	US-60-389-987-2231	112.50	117.12	11.24	
/cgn2_5/ptodata/2/paa/PCT_NEW_COMB.pep	PCT-US02-09944-426	112.00	125.39	12.33	
/cgn2_5/ptodata/2/paa/PCT_NEW_COMB.pep	PCT-US02-08253-242	112.00	124.35	12.30	
/cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep	US-10-155-881-9543	111.50	133.81	13.69	

```
seq_documentation_block:
; Sequence 252, Application US/09994404
; GENERAL INFORMATION:
; APPLICANT: ESSER, KLAUS M.
; APPLICANT: CHAN, JOHN Y.
; APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN
; APPLICANT: DELVECCHIO, ALFRED MICHAEL
; APPLICANT: DILLON, SUSAN B.
; APPLICANT: LEARY, JEFFREY JOSEPH
; APPLICANT: SUTTON, DAVID
; TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES
; FILE REFERENCE: P50583
; CURRENT APPLICATION NUMBER: US/09/994,404
; CURRENT FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 09/297,477
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/030,279
; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: US 60/049,018
; PRIOR FILING DATE: 1997-06-09
; NUMBER OF SEQ ID NOS: 303
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 252
; LENGTH: 3051
; TYPE: PRT
; ORGANISM: Herpes simplex
US-09-994-404-252

alignment_scores:
  Quality: 129.00      Length: 242
  Ratio: 1.372        Gaps: 13
  Percent Similarity: 38.843  Percent Identity: 26.860

alignment_block:
US-09-462-480-3/rev x US-09-994-404-252  ..
Align seg 1/1 to: US-09-994-404-252 from: 1 to: 3051
469 CCCGGTGGCGGGAAGTCTTGTTCATTACGGAGCTCACCATCGTCC 420
|||||  ||||||  |||  |||  |||  |||  |||  |||  |||
2593 ProAlaPheSerGlySerAlaAlaPheSerAlaAlaValProArgVa 2609
419 TCTTCGTCCTCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 370
|||  |||  |||  |||  |||  |||  |||  |||  |||
2609 lArgArgSerArgThrArgAlaLysSerArgAlaProAlaSerA 2626
369 TGCCG..... 365
2626 laProGluGlyTrpArgProProAlaLeuProAlaProValAlaPro 2642
364 .....GCGCGACCA..... 356
2643 ValAlaAlaSerAlaArgProProAspGlnProProThrProGluSerAl 2659
355 .GACCGGGTGGTGGAGCCCGCGGATTGCGAACCTGGCCCATCG.... 311
|||||  |||  |||  |||  |||  |||  |||  |||  |||
2659 aProProAlaTrpValSerAlaLeuProLeuProProGlyProAlaSerA 2676
310 ..... 296
2676 rgAlaPheProAlaProThrLeuAlaProIleProProProAlaGlu 2692
295 .....GAGCGGCGCAC..... 284
2693 GlyAlaValAlaProGlyAspAspArgArgGlyArgArgGlnThrTh 2709
283 .....CCGTACCGACGATCCGCGCAACAGCCCGCGCA..... 251
2709 rAlaGlyProSerProThrProProArgGlyProAlaAlaGlyProProA 2726
251 ..... 251
```

```
2726 rgArgLeuTrpAlaValAlaAlaSerLeuSerAlaSerLeuAsnSerLeuPro 2742
250 TCACCGAGGGGGCAACGGCTTTTCGATCATGAGACATCAGCGCGGTG 201
|||||  |||  |||  |||  |||  |||  |||  |||  |||
2743 SerProArg...AspProAlaAspHisAlaAlaAlaValSerAlaAlaAl 2758
200 CGGTCAACGACCCACTGCCCGAGT.....AGCGACTCCGCGCG 160
|||||  |||  |||  |||  |||  |||  |||  |||  |||
2758 aAlaAlaVal..ProProSerProGlyLeuAlaProThrSerAlaVa 2774
159 CAGCAGGCGCGCGCGCGCTGGGCGCTGATCCA.....C 125
|||  |||  |||  |||  |||  |||  |||  |||  |||
2774 lGlnThrSerProProProLeuAlaProGlyProValAlaProSerGluP 2791
124 CAGCAGCGGGATGTTCGACAGCGGACTGGTCCGAGCAGCGCCCATCTGC 75
|||  |||  |||  |||  |||  |||  |||  |||  |||
2791 rOLeuCysGlyTrp.....ValValPro.GlyGlyProValAl 2803
74 CGGCTTCTCTCGTCTCGGTGGTTCGCCGCGCGGTGCGCGCCACCTGCT 25
||||  |||  |||  |||  |||  |||  |||  |||  |||
2803 aArgArgPro.....ProProGlnS 2810
24 GAACAACGACGTCACCTCTGCA 2
|||||  |||  |||  |||  |||  |||  |||  |||  |||
2810 erProAlaThrLysProAlaAla 2817
seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-22036
seq_documentation_block:
; Sequence 22036, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovacic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 22036
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Glycine max
US-10-155-881-22036

alignment_scores:
  Quality: 127.50      Length: 181
  Ratio: 1.401        Gaps: 11
  Percent Similarity: 50.276  Percent Identity: 35.359

alignment_block:
US-09-462-480-3/rev x US-10-155-881-22036  ..
Align seg 1/1 to: US-10-155-881-22036 from: 1 to: 261
476 CTCCTCCGCGCGGTGGCGGGAAGTCTGTTCATTACGGAGGTCAACA 427
|||||  |||  |||  |||  |||  |||  |||  |||  |||
65 LeuProArgThrAlaThrAlaArgSerThrAlaAlaGlyAlaAlaPheAl 81
426 GTCGTCTCTCTCGTCCAGTCGTCTCTCTCTCTCTCTCTCTCTCTCT 377
|||||  |||  |||  |||  |||  |||  |||  |||  |||
81 aCysArgSerSerAlaProProAlaProAlaSerSerSerArgAlaSerA 98
376 CGACCGGTGGCG...GCGCGACACAGCCGCGCTGGTGGAGCCCGCGGAT 330
|||||  |||  |||  |||  |||  |||  |||  |||  |||
98 laThrSer.ProThrAlaArgProSerSerGlySerSerAlaArgArg.. 113
329 TGCGAACCTTGGGCCATCG...CTCCGCGGACCCACCGAGCGGCCACC 283
```

```
114 ....AsnProProSerProLeuProAlaProAlaProLeuArgA 129
||||| ||| ||||| ||| |||||
282 CGTCACGACGACCGCGCAACAGCCGCGCATCACCGAGGGGCAACCG 233
||||| ||||| ||||| ||||| ||||| ||||| |||||
129 laSerProLeu.....SerProThrThrSerProSerLeuPro 142
||||| ||||| ||||| ||||| ||||| ||||| |||||
232 GCTTTTCGATCAGTACGATCAGCGCGGTGGGTCAACGACCCACCT 183
||||| ||||| ||||| ||||| ||||| ||||| |||||
143 LeuProProSerSerAlaAsnAlaPheAlaLeuThrThr...Pro 158
||||| ||||| ||||| ||||| ||||| ||||| |||||
182 GCGCCAGTAGCGACTCGCGCGCAGAGGCC.....GCGCCCGC 142
||||| ||||| ||||| ||||| ||||| ||||| |||||
159 ProProLysThrThrProSerArgSerProProProArgArgProPr 175
||||| ||||| ||||| ||||| ||||| ||||| |||||
141 GCTGGGCGGTATCCACCGCGCGGTGTCGACAGCGGAGTGTGC 92
||||| ||||| ||||| ||||| ||||| ||||| |||||
175 oSerGlyProLysLeuProAlaProThrSerAlaArgSerGlyAlaSerP 192
||||| ||||| ||||| ||||| ||||| ||||| |||||
91 CGAGCAGGCC...ATCTGCGCGGTCTCTCGTGGGTGGGTGC..... 50
||||| ||||| ||||| ||||| ||||| ||||| |||||
192 roProArgProAsnSerCys.ArgCysArgArgileGlnCysPheTh 208
||||| ||||| ||||| ||||| ||||| ||||| |||||
49 .....GCGCCCGGTGCGCCCGCCT..... 29
||||| ||||| ||||| ||||| ||||| ||||| |||||
208 rIleThrIleAsnThrThrSerSerAsnSerAsnArgProProTrpGluA 225
||||| ||||| ||||| ||||| ||||| ||||| |||||
28 .....GGCTGACACACGACGTCACCTGCT 5
||||| ||||| ||||| ||||| ||||| ||||| |||||
225 rgProLeuProGlnGlySerGlyIleThrSerProAla 237
```

seq_name: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-382-898-278

```
seq_documentation_block:
; Sequence 278, Application US/60382898
; GENERAL INFORMATION:
; APPLICANT: Hudson, Keith
; APPLICANT: et al.
; FILE OF INVENTION: Plant Receptors and Ligands
; CURRENT APPLICATION NUMBER: US/60/382,898
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 1344
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 278
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-60-382-898-278
```

alignment_scores:
Quality: 125.50 Length: 175
Ratio: 1.793 Gaps: 7
Percent Similarity: 40.000 Percent Identity: 29.143

alignment_block:
US-09-462-480-3/rev x US-60-382-898-278 ..

Align seg 1/1 to: US-60-382-898-278 from: 1 to: 708

```
421 CCTTTGTCCTCCAGTCCTCTCTCGTCTTCTTTCACGCTCTCTCGCGAGC 372
||||| ||||| ||||| ||||| ||||| ||||| |||||
38 ProLeuProProSerAlaPro.....ProProAsnArgAl 49
||||| ||||| ||||| ||||| ||||| ||||| |||||
371 GGTGCGCGCGCGACGACCGCGGTGGTGGAGCCGCGCGGATTCGGAAC. 323
||||| ||||| ||||| ||||| ||||| ||||| |||||
49 aProProProProProValThrThrSerProProProValAlaAsnG 66
||||| ||||| ||||| ||||| ||||| ||||| |||||
322 .....CCTGCGCGCTCTCTCGGACCCCGCGCGCGCGCCAC..... 284
||||| ||||| ||||| ||||| ||||| ||||| |||||
66 lyAlaProProProProLeuProLysProProGluSerSerProPro 82
```

```
283 CCCTACCGACGATCCGCGCAACAGCCCGCGCA...TCACCGAGGGGCA 237
||||| ||||| ||||| ||||| ||||| ||||| |||||
83 ProGlnProValIleProSerProProProThrSerProProGln 99
||||| ||||| ||||| ||||| ||||| ||||| |||||
236 ACCGGCTTTTCGATCAGTACGATCAGACATCAGCGCGCTG..... 201
||||| ||||| ||||| ||||| ||||| ||||| |||||
99 nProValIleProSerProProProSerAlaSerProProAlaLeuV 116
||||| ||||| ||||| ||||| ||||| ||||| |||||
201 ..... 201
116 alProProLeuProSerSerProProProProAlaSerValProPro 132
||||| ||||| ||||| ||||| ||||| ||||| |||||
201 ..... 201
133 ArgProSerProSerProProIleLeuValArgSerProProProSerVa 149
||||| ||||| ||||| ||||| ||||| ||||| |||||
200 ....CGGGTCAACAGCACCCAGTGGCGCAGGTAGCGATCCGCGCGCAGCA 155
||||| ||||| ||||| ||||| ||||| ||||| |||||
149 lArgProIleGlnSerProProProProSerAspArgProThrGlnS 166
||||| ||||| ||||| ||||| ||||| ||||| |||||
154 GCGCCGCGCGCGCTGGCGCTGATCCACGACGCGGATGTTGAC 105
||||| ||||| ||||| ||||| ||||| ||||| |||||
166 erProProProProSerProProSerGluArgProThrGln 182
||||| ||||| ||||| ||||| ||||| ||||| |||||
104 AGCGGACTGTGTCGCGAGCGCCATCTGCGCGCTTCTCTCGTGGCTGG 55
||||| ||||| ||||| ||||| ||||| ||||| |||||
183 Ser.....ProProSerProProSerGluArgProThrGlnSerPro.. 196
||||| ||||| ||||| ||||| ||||| ||||| |||||
54 GTTGCCCGCGCGGTGCGCGCCACC 30
||||| ||||| ||||| ||||| ||||| ||||| |||||
197 ....ProProSerProProSer 203
```

seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-25973

```
seq_documentation_block:
; Sequence 25973, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 25973
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-155-881-25973
```

alignment_scores:
Quality: 125.00 Length: 187
Ratio: 1.562 Gaps: 11
Percent Similarity: 42.781 Percent Identity: 29.947

alignment_block:
US-09-462-480-3 x US-10-155-881-25973 ..

Align seg 1/1 to: US-10-155-881-25973 from: 1 to: 318

```
12 GAGCTGCTGTTTCAGCCAGGTGGCGCGCGCGCGCAACCCAGCG 61
||||| ||||| ||||| ||||| ||||| ||||| |||||
120 AspAlaPheProArgAlaGlyAspArgGlnArgArgHis.....Ar 134
||||| ||||| ||||| ||||| ||||| ||||| |||||
62 ACAGGAGAGCGCGCAGATGGCGCTGCTCGCACCAGTCGCTCGCAAC 111
||||| ||||| ||||| ||||| ||||| ||||| |||||
134 gArgGly.....HisArgGlyLeuLeuProp 143
```



```

; APPLICANT: AMSHEY, Stefan R.
; APPLICANT: DAUGHERTY, Sean C.
; APPLICANT: DAM, Tam C.
; APPLICANT: LIU, Tommy F.
; APPLICANT: NGUYEN, Duy-Viet An
; APPLICANT: KLERFELD, Yael
; APPLICANT: GERSTIN JR., Edward H.
; APPLICANT: PERALTA, Careyana H.
; APPLICANT: DAVID, Marie H.
; APPLICANT: LEWIS, Samantha A.
; APPLICANT: CHEN, Alice J.
; APPLICANT: PANZER, Scott R.
; APPLICANT: HARRIS, Bernard
; APPLICANT: FLORES, Vincent
; APPLICANT: MARWAHA, Rakesh
; APPLICANT: LO, Audrey
; APPLICANT: LAN, Ruth Y.
; APPLICANT: URASHKA, Michael
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PT-1232 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09921
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/280,067; 60/291,280; 60/291,849;
; 60/291,829; 60/299,428; 60/300,001; 60/299,776
; PRIOR FILING DATE: 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17;
; 2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20
; NUMBER OF SEQ ID NOS: 1146
; SOFTWARE: PERL Program
; SEQ ID NO 1022
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: LG:007916.8.orf2:2001JUN22
PCT-US02-09921-1022
```

```

alignment_scores:
  Quality: 121.50      Length: 167
  Ratio: 1.293        Gaps: 10
  Percent Similarity: 56.287      Percent Identity: 34.132
```

```

alignment_block:
US-09-462-480-3/rev x PCT-US02-09921-1022 ..
```

Align seg 1/1 to: PCT-US02-09921-1022 from: 1 to: 380

```

479 AGTCTTCGCGCGGGTGGCGGGAAGCTCTGTTGTCATTACGGGAGCTCA 430
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
137 SerLeuTyrAlaAspSerAspGlyGluAsnAspSerThrSerSerSergI 153
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 CCAGTCGCTCTTCGTCGCCAGTCGCTCGTCGCTCTTCTTTCACGCTCCT 380
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 uSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSers 170
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
379 GCGGAGCGGTGCGGCGGACACCGCGGCTGTGGAGCCCGCCGGAT 330
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 erSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSers 186
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
329 TGGCAACCTTGGCCCATCGTCCCG .....GACCCACCGAGCGGGGCC 286
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
187 GluGlu.ArgProAlaAlaLeuProSerAlaSerProProArgGluV 203
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
285 ACCCGTCACCGACG .....ATCCGGCAACAGCGC 257
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
203 alProValProThrProAlaProValGluValProValProGluArgVal 219
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
256 CCGGCATCACGAGGGGCAACCGCTTTTGATCAGTCAGATCAGATCAGC 207
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
220 AlaGlySerProValThrProProAlaArgThrGlyGlyValSerSerLy 236
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
206 GCGGTGC .....GGGTCAAGCACCCACCTCGCGCCAGGTACCGACTC 166
```

```

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 sAlaCysArgProHisGlyValThrProGlnCysAlaSerAlaSerP 253
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165 CGCGCGCGACGA .....GGCCGCGCGCGCTGGGCTGATCCACGAG 122
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
253 roArgThrThrCysTrpAlaProArgPro .....LeuProHisAla 266
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121 CCAGCGGATGTTGCAGACGGGACTGTGGCGGACGAGCCCATCTGCGCG 72
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
267 ProMet .....SerValPro..LeuLeuProSer...ProSerCysPro 279
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71 GCTTCCTCTCGTGGCTGGG .....TTGCCGCGCGCGTGGCGGCCACTGG 27
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 HisProArgAsnAlaGlyLysLeuSerProSerLeuProSerArgTrp 295
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: /cgn_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-360-039-3962

seq_documentation_block:
; Sequence 3962, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3962
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-60-360-039-3962
```

```

alignment_scores:
  Quality: 121.50      Length: 189
  Ratio: 1.538        Gaps: 12
  Percent Similarity: 41.799      Percent Identity: 31.217
```

```

alignment_block:
US-09-462-480-3/rev x US-60-360-039-3962 ..
```

Align seg 1/1 to: US-60-360-039-3962 from: 1 to: 639

```

473 CGCGCCCGGGTGGCGGGAAGTCTGTTGTCATTACGGGAGCTCACCAGTC 424
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
319 ProAlaArgArgSerGlyLys .....LeuAspThrGluAsnHisGln.G 333
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
423 GTCCTCTTCGTCGCCAGTCGT ..... 404
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
333 luProAlaProProProArgPheAlaValProProIleAlaAspAla 349
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 .....CCTCGTGTCTTCTTACGCTCCTGC 378
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 GlyLysPheAlaHisSerAspProProArgHisThrProSerAlaProGl 366
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
377 GCGAGCGGTGCGCGCGGACCA .....GACC 352
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
366 yProProProProArgProProLysThrProLeuGluAspGlnAspP 383
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 CGGCTGTGGAGCGCGGATTCGGAACCTGCCCA ..... 314
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
383 ro .....SerGlnArgPheSerValProProPheThrGlyCln 396
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
313 ...TCGCTCCCGGACCCACCGAGCGGCCAC .....CGGTCCAC 276
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397 ArgSerValProProProProSerArgSerValProProPr 413
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
275 GACGATCCGGCAA .....CAGCGCGCGGCATCACCGAGGGGCAAC 235
```

```

| | | : : | | | | | | | | | | : : | |
413 oProProArgAsnSerAlaAlaGlnProProLeuProProLys...AlaPr 429

234 CGGCTTTTCGATCAGTGCAGACATCAGCGGGTGCGGGTCAACGACCAC 185
|||
429 oGly.....P 431

184 CTCGCCAGTAGGACTCCGCGGCAGCAGCGCCGCCGCGGTG... 138
||||| ||| ||||| ||| |||
431 roAlaProProLeuProProAlaSerSerArgProProMetLeuPro 447

137 ....GGGCTGATCCACAGCACGCGGATGGTTCGACGGGAGTGGT 94
: : : : | | | | : : : : : : : :
448 ThrArgSerProAlaProProGlnAlaProProLeuProThrSerAsnAl 464

93 GCCGAGCAGGCCCATCGCGGGTTCCTCGTGGCTGGGTGCGCGCGC 44
||| ||| ||| ||| : : : : : | | | | |
464 apProProProProLeuProAla...ThrlAlaProPropoProp 480

43 CGGTGCCGCCACC 30
480 roLeuProAlaThr 484
```

seq_name: /cqn2_6/ptodata/2/paa/US09_NEW_COMB pep: US-09-784-984A-8

```

seq_documentation_block:
; Sequence 8, Application: US/097844984A
; GENERAL INFORMATION:
; APPLICANT: AUDONNET et al.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY PATHOLOGIES
; FILE REFERENCE: 454313-2230
; CURRENT APPLICATION NUMBER: US/09/784,984A
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US/09/232,468A
; PRIOR FILING DATE: 1999-01-05
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Aujeszky's Disease Virus (NIA3 Strain)
US-09-784-984A-8

```

alignment_scores:		
Quality:	121.00	Length: 166
Ratio:	1.424	Gaps: 7
Percent Similarity:	51.205	Percent Identity: 30.723

```

alignment_block:
US-09-462-480-3/rev x US-09-784-984A-8      ..
Align seg 1/1 to: US-09-784-984A-8 from: 1 to: 404

```

472 CGGCCCGGGTGGCCGGGAGTCTGTTGCATTACGGGAG.....CTCACGAC 426
 193 ArgProArgGlyProAlaProHisValGlnValArgValLeuGluAr 209
 425 TCGTCTCTTCGTCCCACTCGCTCTCTCGTCGTCTTCTTCCAGTCCTGGCG 376
 209 gArgGlnLeuGlnAlaGlyArgGlyArgAspAlaIle.....ProAspA 224
 375 GAGCGGTGCGGGCGCGACACACCGCGGCTGTGTGAGCGCGCGGATTCGC 326
 224 laValLeuProAlaAlaProAlaProGlyGlyGlyGluLeuValPro 240
 325 AACCTTGCCCATCGTCCCGGACCCACCGAGCGGGCGCACCCTGCAC 276
 241 GlnGluArgProAspAlaProAlaGlyProArgArgHis..... 254
 275 GACGATCGGCAACAGCGCGCGGA.....TCAC 247

255 AlaValArgHisArgProArgAlaAlaLeuGlyClyLeuAlaGluAlaP 271
246 CGAGGGGCAACCGGCTTTTCATCAGCTGAGCATCAGCGCGCTCGGG 197
|| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
271 roAlaProAlaProAlaProAlaProAlaGluAlaArgAlaArgProGly 287
196 TC AACGCCACTGCGCCAGGTAGCNACTCCGCCGCAGCAGGCCCGG 147
|| : ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
288 AspAlaArgAlaProArgProProAlaArgAlaGlyAspAlaGlyProAr 304
146 CCGCGCTGGGCGCTGATCCACCAGCCAGCGGATGGTTCAGCACGGGACT 97
|| | ||| ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
304 gArgArg.....GlyProProHisAlaAlaThrProGluA 316
96 GGTGCCGACGAGC.....CCATCTGGCGGGCTTCCCTCGTCGG 59
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
316 laArgAspAlaAlaProProLeuArgProAlaGlyArgAlaGlnArg 332

58 CTGGTGTCGCGCGCGTCCGCCCCACCTGGCTGAACAACGACGT 13
333 ValAlaAlaAlaArgGlyAlaValProAlaAlaAlaAspProArgArg 347

seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-11183

```

seq_documentation_block:
; Sequence 11183, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotsen, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lufiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; FILE OF INVENTION: TRANSCRIPTION IN
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 11183
; LENGTH: 668
; TYPE: PRT
; ORGANISM: zea mays
US-10-155-881-11183

```

alignment_scores:		
Quality:	121.00	Length: 178
Ratio:	1.344	Gaps: 8
Percent Similarity:	50.562	Percent Identity: 29.775

```
alignment_block:
US-09-462-480-3/rev x US-10-155-881-11183  ..
Align seg 1/1 to: US-10-155-881-11183 from: 1 to: 668
```

481	CAAGCTCTCCGGCCCGGGTGC	CGCGGGAAGCTCTGTGTGTCAT	TACGGGAGCT	432
:				
171	GlutIleHisProCysAsnProGly	..LeuAlaSerLeuLeuIle	186	
431	CACCAAGTCGTCTCTTCGTCCCA	409	
:				
186	uArgLeuArgSerLeuLeuAlaAla	AspProAlaProLeuProPro	203	
408	GTCTGCTCTCGNCTCTCTCTCACGCTCTCGCGAGCGGTGCGG	365	
:				
203	roGlnProGlnHisAlaLeuLeuHisGly	AlaProAlaAlaPro	219	
364	CGCCSACACAGCCGGGTGTGTGACCGCGCGATTTCGAA	324	
:				
220	AlaGlyLeuThrLeuProPro	227	
323	CCCTGGCCCATCGTCTCCGGACCCACCGAGCGGCGCCACCCGTCACCGA	274		
:				

```

alignment_scores:
  Quality: 120.50      Length: 233
  Ratio: 1.242        Gaps: 11
  Percent Similarity: 41.631  Percent Identity: 25.751

alignment_block:
  US-09-462-480-3/rev x US-10-155-881-10047  ..

Align seg 1/1 to: US-10-155-881-10047 from: 1 to: 332

479 AGTCTTCGGCGCGGTGGCGCGG.....AAGTC 451
      ||| |||||:||||: |||
34 SerAlaProSerArgThrArgGlyArgArgGlyAlaAlaThrArgAl 50

450 TGTGTCTATACGGAGCGCTCACGAGTGTCTCTCGTCCCGAGTCGTCT 401
      : : : ||| ||| : : : |||||:||||: |||
50 aSerThrThrProAlaProAlaArgThrSerAlaArgAlaGlyA 67

400 CGTGGCTCTTCTCA.....CGTCCGCGCGAGCGGTGCCGCG 363
      :|||||:||||: ||||: |||||:|||||
67 laSerAlaAlaThrAlaAlaThrTrpArgThrProAlaSerGlyAlaGly 83

362 GCGACACGACCGCGGTGGTGAGCCGCGAGTTGCGAACCCCTGGCCCCAT 313
      :||||||| : : : : |||

```

```

alignment_scores:
  Quality: 120.50      Length: 140
  Ratio: 1.746        Gaps: 8
  Percent Similarity: 49.286      Percent Identity: 35.000

alignment_block:
  US-09-462-480-3/rev x US-10-155-881-10704 ..

```

Align seg 1/1 to: US-10-155-881-10704 from: 1 to: 356

```
460 CCGGAAGCTGTGTTGCTATTACGGAGCTCACC...AGTCGCTCTCTTCG 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 ProAlaAlaProLeuProCysArgArgCysThrProThrProThrAr 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
413 TCCAGTGTCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 373
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 gSerThrArgGlnArgArgGlnArgThrProAlaArgThrAlaAla 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 CGGTGCGCGCGCGACAGACCGCGGCTGCTGGAGCGCGGATTCGGA 323
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 laSerProThrArgProProProProProProProProProPro 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 CCGTGGCCATCGCTCCCGGACCCACCGGAGCGG..... 290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 ThrSerProSerLeuProProProProProProProProProPro 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
289 .....CGCCACCGGTACCGAGTCGCGGCAACAGCGCGCGCATCAC 247
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
279 rTrpProArgAlaProArgArgSerHisArgAlaArgAlaProSer 296
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 CGAGGGGCAACCGGCTTTTCAGTCAGCTGAGACATCAGCGCGCTGG 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
296 roArgArgArgArg.....ArgAlaProThrThrArgSerArgCys 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 TCAACGACCCACCTCGCGCAGGTAGCGACTCCGCGCGCAGCGCGCG 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 SerSerThrSerSerAlaSerAlaArgArgAspGlyAlaGlyProGl 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 .CGCCCGCGCTGGCGCTGATCCAGCAGCGGATGTTTCGACAGCGG 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
328 yArgProGlyGlyGlyVal.....GlnGluGlyGlyGlyGlnArg 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
99 ACTGTGTCCGAGCGGCC 82
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 ThrArgAlaProAlaAla 348
```

seq_name: /cgn2_6/ptodata/2/paa/us09_NEW_COMB.pep:US-09-573-655B-274

seq_documentation_block:

```
; Sequence 274, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 274
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-274
```

alignment_scores:
Quality: 119.50 Length: 135
Ratio: 1.757 Gaps: 4
Percent Similarity: 50.370 Percent Identity: 28.889

alignment_block:

US-09-462-480-3/rev x US-09-573-655B-274 ..

Align seg 1/1 to: US-09-573-655B-274 from: 1 to: 312

```
425 TCGTCTCTTCTGTCCTCCAGTCGCTCGTCTCTCTCTCTCTCTCTCT 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 ThrThrSerSerSerSerLeuSerProSerSerSerSerProSerLeu 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
375 GAGCGGTGCCCGCGGACAGACCGCGGCTGGTGGAGCGCGCGGATTG 326
```

```
37 rProSerProProSerSerSerProSerSerAlaProProSerSerLeu 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 AACCTGGCCCATCGCTCCCGGACCCACCGGAGCGCGCGCCGTCACC 276
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 erPro.....SerSerProProProProLeuSer 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
275 GACGATCGGCAACAGCGCGCGCATCACCGAGGGGCAACCGGCTTTC 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63 LeuSerProSer..... 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 GATCAGCTGACATCAGCGCGTGGGGTCAACGACCCACCTGGCCGAC 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 .....SerProProProProProp 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
175 GTAGCGACTCGCGCGCAGCAGCGCGCGCTGGGGCTGATCA 126
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 roSerSerProLeuSerSerLeuSerProSerLeuSerProSerPro 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 CCAGCCAGCGGATGTTTCGAC.....AGCGACTGGTGGCGAGCAG 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 ProSerSerProSerSerAlaProProSerSerLeuSerProSerSe 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 GCCCATCTGCGCGCTTCCTCGTGGTGGTGGCGCGCGCGGTGCGGC 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
106 rPro...ProProLeuSerLeuSerProSerSerProProProProp 122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 CCACC 30
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 roSer 123
```

seq_name: /cgn2_6/ptodata/2/paa/us10_NEW_COMB.pep:US-10-155-881-7219

seq_documentation_block:

```
; Sequence 7219, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 7219
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-7219
```

alignment_scores:
Quality: 119.00 Length: 185
Ratio: 1.266 Gaps: 10
Percent Similarity: 50.811 Percent Identity: 27.568

alignment_block:

US-09-462-480-3/rev x US-10-155-881-7219 ..

Align seg 1/1 to: US-10-155-881-7219 from: 1 to: 307

```
470 GCCGGGTGGCGGGAAGTCTGTTGTCATTACGGGAGCTCACCAGTCGTC 421
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
19 AlaArgAlaProSerSerAlaAlaCysGlyGlyArg.....G1 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 CTCCTTCGTCAGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
33 yAlaSerSerArgArgArgSerAlaThrArgGlyAlaAlaCysAlaSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 GTGCGCGCGGACAGACCC.....GGGCTGGTGAG 339
```

```
|||||: : : |||
50 lySerAlaProSerThrProArgArgProArgAlaSerThrPro 66
338 CCGCGGATTCGAACCGTCCCATCGCTCCG .....G 304
|||||: : : |||
67 ProProSer.SerSerAlaGlyProProProProProSerProArgp 83
|||||: : : |||
303 ACCACCGGCGCGCGCCGTCACCGAGATCCGCAACAGCCGCG 254
|||||: : : |||
83 roProProLeuThrAlaAlaProGlyPro.GlyProGlyArgThrArg 99
253 GCATCAGCGAGGGGCAACCGCTTTCGATCAGCTGAGACATCAGCGGC 204
|||||: : : |||
99 gArgHisArg.....GlnArgAlaThrSerProAlaArgSerArgArg 114
203 GTGGCGGTCAACGAC.....CCACCTGCGCCAGGTAG 172
|||||: : : |||
114 r9.ArgProCysSerAlaArgSerProProCysProProSerProArgTh 130
171 CGACTCCGCGCGCAGCGCCGCGCGCTGGGCGCTGATCCACCG 122
|||||: : : |||
130 rThrProThrArgArgProProSerAlaAlaAlaTrpProSerSerArg 147
121 CCAGCGGATGCTGCAC.....ACGGGACTGGTCCGAGCAGGCC 81
|||||: : : |||
147 r9ArgSerTrpAlaSerSerCysArgSerArgThrArgProSerThrPro 163
80 ATCTCGCGCGCTCTCTCGGCT..... 57
164 ProAlaAlaSerGlyThrSerSerProThrArgGlySerSerThrProSe 180
56 .....GGGTTGCCCGCGCGGTGCCGCCACC 30
180 rArgArgArgArgArgArgArgGlyMetProProProProSerProSer 196
seq_name: /cgu2_6/ptodata/2/paa/US10_NEW_COMB.pep.US-10-155-881-26034
```

```
seq_documentation_block:
; Sequence 26034, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 26034
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-155-881-26034
```

```
alignment_scores:
Quality: 118.50 Length: 180
Ratio: 1.362 Gaps: 7
Percent Similarity: 48.333 Percent Identity: 28.889
alignment_block:
US-09-462-480-3 x US-10-155-881-26034 ..
Align seg 1/1 to: US-10-155-881-26034 from: 1 to: 349
30 GTTGGCGCGCACCGCGCGCGCCAGCCAGCGAGCAAGCAAGCGCGCAGA 79
|||||: : : |||
62 GlyGlyArgGlyGlyGluArgHisValArgArgAla...GlyG1 77
80 TGGGCTCTCTCGGCACCATGTCGTCGAACCATCCGCTGGCTGGTGGGA 129
```

```
||| |||: : : ||| ||| ||| ||| |||
77 yGlyGlyAlaGluGluGluAlaArgArgGlyAlaGlyGlySerTrp 94
130 TCAGGCCCGCGCGCGCGCGCTGCTGCGCGCGAGTCCGTACCTGG 179
|||: : : ||| |||: : : |||
94 r9ArgGlyArgArgGlyGlyGlyAlaGlySerProAlaAlaGlyArg 110
180 CGCAGGTGGTCTGTTACCGCGCGCGCTGATGCTCTCAGTCATCGAAA 229
|||||: : : ||| |||: : : |||
111 ArgArg.....ProProArgAlaProProCysAsnAlaGlyGlyG1 124
230 AGCGGTGTCCTCGCTGATGCGCGCGCTGTTGCGGATCCTCGGT. 278
: |||: : : ||| |||: : : |||
124 u.....AspValGlyGlyAlaCysGlyAlaGluGlyG 135
279 .....GACGGTGGCGCGCTCC.....GTTGGTCCGGG 308
|||||: : : ||| |||: : : |||
135 luAlaArgGluAlaAspGlyGlyHisArgGlyGlyGlyGlyGluAla 151
309 AGCATGGCGCGCGGTTCGCAATCCGCGCGCTCCACCGCGCGGTCTGG 358
: : : ||| |||: : : ||| |||: : : |||
152 AlaGluGlyGlnGlyArgGlyAspArgAlaAspArgSerProHisLeuG1 168
359 TCGC.....GCCGGCACCGCTC 375
168 yAlaArgGlyAlaProLeuGlyHisValArgArgGlyAlaGlyValAla 185
376 GCGCAGGAGCGTGAAGA..... 392
|||: : : |||
185 r9ProArgAlaValGlnArgGlyHisArgGlnArgProProArgArgAla 201
393 .....AGCGACGAGGACGACTGGG 412
|||||: : : |||
202 ArgAlaArgProArgArgProArgAlaProArgArgArgProArgArg 218
413 ACGAAGAGGAGGACTGGTGAGCTCCCGTAATGACAACAGA 452
|||||: : : ||| |||: : : |||
218 gArgArgProArgValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 231
```

OM of: US-09-462-480-3 to: PIR_71:* out_format : pfs
 Date: Jul 22, 2002 1:26 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+nt2p.model -DEV=xlh
 -O=/sqn2.1/USPTO_spool/US09462480/runat_18072002.164418.19445/app_query_fasta_1.2850
 -DB=PIR_71 -QEXT=fastan -SUFFIX=rpr -GAPOP=12.000 -CAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DSCALE=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODEL=LOCAL -OUTEXT=pfs -NORM=ext -HEAPSIZ=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09462480_@CGN1_1_73
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-3
 Query length: 481
 Database: PIR_71:*
 Database sequences: 283138
 Database length: 96089334
 Search time (sec): 87.620000

score_list:

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
pir2:G70802	+	713.00	943.67	4.9e-45	368	! Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
pir2:T10033	+	198.00	264.03	4.3e-07	302	! Hypothetical protein MLCB28.14
pir2:S52996	-	140.00	185.24	0.0078	403	! prpL2 protein - human (fragment)
pir2:S20590	-	133.00	170.07	0.0242	913	! exo-alpha-sialidase (EC 3.2.1.1)
pir2:A46302	+	131.00	169.28	0.0346	707	! PTB-associated splicing factor, elastin precursor - mouse
pir1:RAMS	+	127.00	162.57	0.0672	860	! Hypothetical protein F24J5.8
pir2:D96711	-	125.50	161.99	0.0879	708	! DNA-binding protein SE5 - rat
pir2:JC4163	+	125.50	160.89	0.0869	825	! Hypothetical protein F5088A.9
pir2:T31611	+	125.00	155.51	0.0902	1585	! Hypothetical protein F54B11.2
pir2:T22602	-	124.50	166.78	0.1108	304	! Probable reductase (EC 1.3.99.1)
pir2:T70507	-	124.50	159.08	0.1025	882	! Probable ABC transporter - Streptococcus
pir2:T35192	-	124.00	159.65	0.1129	744	! Hypothetical protein Rv1158c - Mycobacterium tuberculosis
pir2:T70595	-	123.00	166.91	0.1460	227	! exo-alpha-sialidase (EC 3.2.1.1)
pir2:S20590	+	132.50	156.19	0.1435	913	! Chitinase (EC 3.2.1.14) precursor
pir2:S51939	-	121.00	159.50	0.1953	439	! Probable multi-domain regulator of cell growth
pir2:T50568	-	120.50	150.80	0.1959	1334	! Hypothetical protein (sdsB regulon)
pir2:A61183	-	119.50	159.98	0.2582	312	! Infected cell protein ICP34.5 - Herpesvirus
pir1:WMBE38	-	119.00	158.32	0.2782	358	! Hypothetical protein - Deinococcus
pir2:T75457	-	119.00	157.61	0.2762	395	! Hydroxyproline-rich glycoprotein
pir2:S06733	-	119.00	154.36	0.2673	620	! Collagen alpha 2(I) chain - bovine
pir1:GCB025	+	118.50	156.83	0.3003	402	! Abi substrate ena (enabled) - Arabidopsis
pir2:A56154	+	118.50	152.98	0.2888	684	! ABC transporter, ATP-binding protein
pir2:T75311	+	118.00	153.87	0.3194	552	! Extensin homolog F28A21.80 - Arabidopsis
pir2:T40459	-	118.00	150.85	0.3097	839	! Hypothetical protein DKFZ0434A1
pir2:T46289	-	118.00	150.65	0.3091	862	! Unknown protein [imported] - Arabidopsis
pir2:G86441	-	118.00	148.26	0.3017	1201	! Limb deformity (ld) protein - Arabidopsis
pir2:T41724	-	118.00	148.18	0.3015	1213	! Hypothetical protein Rv2082 - Mycobacterium tuberculosis
pir2:T70766	-	117.50	151.28	0.3409	721	! Hypothetical protein - Sulfolobus
pir2:T40505	+	117.50	144.06	0.3169	1958	! Hypothetical protein W04B5.3 - Helicobacter
pir2:T33663	-	117.00	154.46	0.3858	424	! Pistil extensin-like protein - Arabidopsis
pir2:T01696	-	117.00	154.42	0.3857	426	! Immediate-early protein - Sulfolobus
pir1:A45344	-	117.00	145.59	0.3527	1446	! Hypothetical protein F13A7.1 - Arabidopsis
pir2:T20807	+	116.50	155.39	0.4268	340	! Homeotic protein BarH1 - fruit fly
pir2:T339369	-	116.50	152.01	0.4124	543	! Drebrin E2 - chicken
pir2:T43776	-	116.00	150.54	0.4453	607	! Collagen COL1 - freshwater species
pir2:T351521	+	115.50	147.78	0.4745	812	! Extensin-like protein - maize
pir2:T349115	+	115.50	145.03	0.4614	1188	! Probable nuclear antigen - Sulfolobus
pir1:845344	+	115.50	142.30	0.4489	1733	! Proline-rich protein APG - Arabidopsis
pir2:S21961	-	115.00	150.14	0.5325	534	! Wiskott-Aldrich syndrome protein
pir2:T43556	-	115.00	149.62	0.5297	574	

pir2:T38819 - 115.00 149.62 0.5297 574 ! Wiskott-Aldrich syndrome protein
 pir2:S54986 - 115.00 145.76 0.5094 980 ! regulatory protein - Emerine
 pir2:T35474 - 114.50 150.62 0.5864 456 ! 50kD proline rich protein -
 pir2:SL3367 - 114.50 148.57 0.5743 606 ! Om(1D) protein - fruit fly (fi
 pir2:T13078 - 114.50 146.82 0.5642 772 ! KIAA0992 protein - human (fi
 seq_name: pir2:G70802
 seq_documentation_block:
 probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: G70802
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen
 A:Reference number: A70500; MUID:98295987
 A:Accession: G70802
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-368 <COL>
 A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL7965.1; PID:e12
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: PPE

alignment_scores:
 Quality: 713.00 Length: 143
 Ratio: 5.057 Gaps: 0
 Percent Similarity: 98.601 Percent Identity: 96.503
 alignment_block:
 US-09-462-480-3 x G70802
 Align seg 1/1 to: G70802 from: 1 to: 368

1 CTGACGAGTGTGCTGTTTTCAGCCAGTGGGGCCACCGGGGGGGG 50
 |||||
 226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyGly 242
 51 CAACCCAGCCGACGAGGAGCGCGAGATGGCTGCTGGCCACCGAGTC 100
 |||||
 242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
 |||||
 101 CGCTGTCTGAACCATTCGCTGGCTGGATCAGGCCCGCCAGCGGGCGCG 150
 |||||
 259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
 |||||
 151 GGCCTGTCTGCGGGAGTGGCTACCTGGCGGAGTGGGTGGTTCACCCG 200
 |||||
 276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
 |||||
 201 CACGCCGTGTATGCTCAGCTGATCTGAAAGCCGTTGCCCGCTCGGTGA 250
 |||||
 292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
 |||||
 251 TGCCGGCGGCTGTGTCGCCGATTCGCTGGTGACGGGTGGCGCGCTCCCGTG 300
 |||||
 309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
 |||||
 301 GGTCCGGGAGGCGTGGCCAGGGTTCGCAATCCCGGGCTCCACCGCCC 350
 |||||
 326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
 |||||
 351 GGGTCTGTCTGCGCGCGCCAGCGCTGCGGAGGCGTGAAGAAGACGACG 400
 |||||
 342 oglyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspaspG 359
 |||||
 401 AGGACGACTGGGACGAAGAGGACGACTGG 429

A:Residues: 1-913 <HEI>
A:Cross-references: EMBL:X62276; NID:g39254; PIDN:CAA44166.1; PID:g39255
A:Reference number: A46302; MUID:93194059
C:Keywords: glycosidase; hydrolase

alignment_scores:
Quality: 133.00 Length: 188
Ratio: 1.529 Gaps: 9
Percent Similarity: 46.277 Percent Identity: 29.787

alignment_block:

US-09-462-480-3/rev x S20590 ..

Align seg 1/1 to: S20590 from: 1 to: 913

```
472 CGCGCGGGTGGCGGGAAGTCTGTCTCATTACGGGAGCTCACCAGTCG 423
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
586 ArgProArgArgProArgArgAlaLeuSerProArgArgHisArgHisH1 702
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 TCCTCTTCGTCGCAGTGGCTCGTCTGCTCTTCTTACGGCTCCGCGGAG 373
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
702 sProProArgProSerArgAlaLeuArg.....ProSerArgA 715
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 CGGTGCGCGG.....CGACCAGACCGCGGCTGGTGAGC 338
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
715 laGlyProGlyAlaGlyAlaHisAspArgSerGluHisGlyAlaHisThr 731
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 CGCGGATTGCGACCTGCGCCATCGCTCCCGGACCCACCGAGCGGCG 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
732 GlySerCysAlaGlnSerAlaProGluGlnThrAspGlyProThrAlaA1 748
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 CCACCCGTCACCGACGTCGCGCAACAGCGCGGCATCACCGAGGGGCG 238
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
748 a....ProAlaProGluThrSerSerAlaProAlaAlaGluProThrClnA 764
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 AACGGGCTTTTGA.....TCAGCTGAG 215
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
764 laProThrValAlaProSerValGluProThrGlnAlaProGlyAlaGln 780
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 ACATCAGCGCG.....TCGGGTCAACGACCCACCTGCGCCAGGTAGC 171
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
781 ProSerSerAlaProLysProGlyAlaThrGlyArgAlaProSerValva 797
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
170 GACTCCGCGCGACGAGC..... 152
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
797 lAsnProLysAlaThrGlyAlaAlaThrGluProGlyThrProSerSers 814
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 .....CGCGCGCGCGCTGGGC 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
814 erAlaSerProAlaProSerArgAsnAlaAlaProThrProLysProGly 830
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
133 CTT.....ATCCACCGCCAGCGGTGTTTCGACAGCGGACTG 96
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
831 MetGluProAspGluIleAspArgProSerAspGlyThrMetAlaGlnPr 847
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 GTGCCGAGCAGGCCATCTCGCGGCTTCCTCGTCGGCTGGTTCGCCGCC 46
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
847 oThrGlyAla....ProAlaArgArgValProArgArg.....ArgA 860
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45 GCGGGTCCGCCCA 32
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
860 rgArgArgArgPro 864
```

seq_name: pir2:A46302

seq_documentation_block:

PTB-associated splicing factor, long form - human
N:Alternate names: myoblast cell surface antigen 24.1D5; polypyrimidine tract-binding pr
C:Species: Homo sapiens (man)
C:Date: 21-Sep-1993 #sequence revision 18-Nov-1994 #text_change 24-Sep-1999
C:Accession: A46302; A43557; S29995
R:Patton, J.G.; Porro, E.B.; Gaiceran, J.; Tempst, P.; Nadal-Ginard, B.
Genes Dev. 7, 393-406, 1993

A:Title: Cloning and characterization of PSF, a novel pre-mRNA splicing factor.
A:Reference number: A46302; MUID:93194059
A:Accession: A46302
A:Molecule type: mRNA
A:Residues: 1-707 <PAT>
A:Cross-references: EMBL:X70944; NID:g38457; PIDN:CAA50283.1; PID:g38458
A:Note: Sequence extracted from NCBI backbone (NCBIF:127206)
R:Gover, H.J.; Moore, S.E.; Dickson, G.; Elsom, V.L.; Nayak, R.; Walsh, F.S.
Development 105, 723-731, 1989
A:Title: Cloning and characterization of a myoblast cell surface antigen defined by 2
A:Reference number: A43557; MUID:90091812
A:Accession: A43557
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 312-707 <GOW>
A:Cross-references: GB:X16850; NID:g23711; PIDN:CAA34747.1; PID:g23712
C:Genetics: A:Gene: GDB:SFPQ; PSF
A:Cross-references: GDB:138275
A:Map position: 4q-4q
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleopro
C:Keywords: alternative splicing; pre-mRNA splicing; surface antigen
F:298-359/Domain: ribonucleoprotein repeat homology <RRM1>
F:372-438/Domain: ribonucleoprotein repeat homology <RRM2>

alignment_scores:

Quality: 131.00 Length: 139
Ratio: 1.899 Gaps: 9
Percent Similarity: 49.640 Percent Identity: 31.655

alignment_block:

US-09-462-480-3/rev x A46302 ..

Align seg 1/1 to: A46302 from: 1 to: 707

```
367 CGCGCGGACGACGCGGGCTGTGGAGCGCGGATCGCAACCTGG 318
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
46 ProMetGlyProGlyProGlyGln.....SerGlyProLy 57
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 CCATCGCTCCGACCCACCG.....GACGCGCGCCACCGTCAC 277
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 sProProileProProProProHisGlnGlnGlnGlnProProp 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 CGACGATCCGCAACACGCGCGGCATCA.....CCGAGGGGCAACCG 233
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
74 ro.....GlnGlnProProGlnGlnProProHisGlnPro 87
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 GCTTTTCGATCAGTACATCAGCGGCTGCGGTCAAGACCCACCT 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
88 ProProHisProGlnPro.HisGlnGln.....GlnGln.ProPro 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
182 GCGCCAGGTAGGACTCCGCGCG.....AGCAGGCC 151
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 ProProProGlnAspSerSerLysProValAlaGlnGlyProGlyPr 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
150 CGCGCGCGCTGGGCTGATCCACGACGCGCGGA..... 114
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 oAlaProGlyValGlySerAlaProAlaSerSerSerAlaProProA 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
113 .....TGGTTCGACAGCGGACTGGTG 93
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 laThrProThrSerGlyAlaProProGlySerGlyProGlyProThr 150
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
92 CCGAGCAGGCCCATCTCGCGGGTTCCTCGTCGGCTGGTTCGCCGCC 43
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ProThrProProAlaValThrSerAlaProProGlyAlaProProPr 167
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
42 GGTGCCGCCACC 30
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 oThrProProSer 171
```

seq_name: pir1:EAMS

174 AspGlyVG¹¹AlaGlySerProAlaAlaProSerProProGly¹⁸⁷

242 GGGGCAACCGGCTTTTCGATCAGCTGAGACATCAGCGGCGTCGGGGTCAA 193

242 GGGGCAACCGGCTTTTCGATCAGCTGAGACATCAGCGGCGTCGGGGTCAA 193

815 GlyAlaLysArg.....ProGlyAlaLysLysTh 824
192 CCACCCAGCTGGCGAGTAGGACTCCGCGCGCAGCAGCGCCGCGCG 143
: ||||| ||||| : : : : : |||||
824 rProProAlaProGlyLeuAlaGluProAlaAlaGlnProGlnProG 841
142 CGCTGGGGCTGATCCACGACCCAGCG.....GA 114
: ||||| : : : : : |||||
841 luAlaLysProGlnPro.GluProAlaAlaProProLysProGlnThAs 857
||||| : : : : : |||||
113 TGGTTCGACAGCGGACTGTCGCGAGCAGGCCATCTCGCGGCTTCCTC 64
857 polyAspPro.....AlaAlaProAlaProVal...L 868
63 CTCGCTGGTGGTGGCGCGCGGTGGCGCCACCTGGC 26
: : ||||| || : : ||||| |||||
868 ysglyLeuGlylleAlaargGlyAlaArgProProGly 880

seq_name: pir2:T35192

seq_documentation_block:

probable ABC transporter - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T35192

R:Seeger, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

A:Reference number: 221571

A:Accession: T35192

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-744 <SEE>

A:Cross-references: EMBL:AL022374; PIDN:CAA18516.1; GSPDB:GN00070; SCOEDB:SC5B8.08

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC5B8.08

alignment_scores:
Quality: 124.00 Length: 142
Ratio: 1.610 Gaps: 6
Percent Similarity: 54.225 Percent Identity: 31.690

alignment_block:
US-09-462-480-3/rev x T35192 ..

Align seg 1/1 to: T35192 from: 1 to: 744

437 GGAGTCCACAGTCGTCCTCTTCGTCGCCAGTCGTCGTCGTCCTTC 388
||||| : : : : : |||||
293 GlyAlaGlyGluIleProAlaAlaArgGlnGlnGluArgArgGluAlaAs 309
387 AGGCTCTCGCGGACGGTGGCGG.....CGCAGCAGACCGCGG 347
: : : ||| ||||| ||| |||||
309 pGlyGluArgAlaProGlyThrGlyThrLeuSerAlaGlyArgProGly. 325
346 TGGTGGCGCGCGGATGCGAACCCCTGCGCCATCGCTCCCGACCCACC 297
||||| : : : ||| : : : |||||
326ProGlu...ThrProGlyValLeuProSerProAla. 337
296 GGAGCGCGCCACCGTCACCGACGATCCGGCAACAGCGCGCGGATCAC 247
: : : ||| ||| : : : |||
338 ThrGluGlyGluProLeuGluAspGlyArgGlyAlaGlyHisGlyAspG 354
246 CGAGGGGCAACCGGCTTTTCGATCGCTGAGACATCAGCGCGCTCGGG 197
: : : ||| : : : ||| : : : |||
354 yAspGlyProAlaSerValAlaValProAsnArgThrProGlyProArg. 370
196 TCAACGACCACTCGGCA.....GGTAGGACTCCGCGCGC 159
: : : ||||| ||| : : : |||||
371GlnAlaProAlaProValSerGlyHisGlyProGluAlaAlaPro 385
158 AGCAGCGCGCGCGCTGGGCGCTGATCCACGACCGCGATGTT 109

||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
386 SerProSerAlaProAlaProGlyProSerGluProAlaSerGly..... 400
108 CGACAGCGGACTGGTGGCAGCAGCGCCATCGCGCTTCCTCGTGGG 59
||| : : : : : ||||| : : : : : |||||
401 .ProSerAlaProAlaProGlyProProAlaProAlaGlyProSerA 417
58 CTGGGTTGGCGCGCGGTGGCGGCC 33
||| ||| ||| |||
417 laProAlaProGlyProSerAlaPro 425

seq_name: pir2:G70555

seq_documentation_block:

hypothetical protein Rv1158c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: G70555

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sqaers, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen

A:Reference number: A70500; MUID:98295987

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-227 <COL>

A:Cross-references: GB:295584; GB:AL123456; NID:g3261774; PIDN:CAB09017.1; PID:g2117

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv1158c

C:Superfamily: proline-rich protein 3

alignment_scores:
Quality: 123.00 Length: 220
Ratio: 1.281 Gaps: 11
Percent Similarity: 43.636 Percent Identity: 26.364

alignment_block:
US-09-462-480-3/rev x G70555 ..

Align seg 1/1 to: G70555 from: 1 to: 227

464 GTGGCGGGAAGTCGTGTCATACGGGA.....GCTCAC..... 429
||||| : : : : : |||||
15 ValGlySerSerAlaAlaLeuLeuThrGlyIleAlaHisAlaAspPr 31
428CAGTCGCTCTCTCT 413
||||| : : : : : |||||
31 oAlaProAlaProAlaProAlaProAsnIleProGlnGlnLeuIleSerS 48
412 CCCAGTCGTCCTCGCTGCTTCTTCACGCTCTCGCGAGCGGTCGCGG 363
|| : : : : : ||| : : : : : |||
48 eAlaAlaAsnAlaProGlnIleLeuGlnAsnLeuAlaThrAlaLeuGly 64
362 GGCACACACCGCGGCTGGTGGAGCGCGGATTCGGAACCTGGCCCAT 313
||||| ||| ||| : : : ||||| |||
65 AlaThr...ProProLeuSerAlaProLysValAlaGluProAlaProAl 80
312 CGCTCCCGGACCCACGAGCGCGCGCTCACCGAC...GATCCGG 266
||||| ||| : : : ||| : : : |||
80 aAlaProGlyIleThrAlaThrPheProGlyLeuThrProAlaAlaProA 97
265 CAACAGCGGC..... 255
||| : : : |||
97 laAlaAlaAlaAlaProAlaLeuThrProSerIleProGlyValAsnAla 113
254GGCACCAGCGGCGGCAACCGGCTTT..... 228
||||| : : : ||| : : : |||
114 ProIleProGlyIleThrProAlaAlaProAlaLeuProValThrAlaPr 130

227TCGATCAGCTGAGACATCAGCGCGTGC 200
 130 oAlaAlaAlaProThrIleProGlyValAsnAlaProIleProGlyIleT 147
 199 GGGTCAACGACCCACCTGG.....CCAGGTAGC 171
 147 hAlaProAlaProAlaAlaAlaValProAlaSerValProGlyVal 163
 170 GACTCCGCGCGC..... 159
 164 ProSerAlaLysValAspLeuProGlnLeuProTyrLeuProLeuGlnVa 180
 158ACGAGCCCGCG.....CCCGGCTGGGGCTGATC 128
 180 lProGlnGlnLeuSerLeuProAlaAspLeuProAlaLeuAla..... 194
 127 CACGAGCCAGCGGTGTCGACAGCGGACTGGTCCGAGCAGGCCATC 78
 195SerGlyValIleProAlaAlaProIle 203
 77 TCGCGGCTTCCTCGCTGGGTGCGG.....CCGCCGTGCCGCC 34
 204 AlaProThrProAlaProGlyAlaProAlaLeuProGlyProPr 220
 33 CACTGGGTG 24
 220 oSerLeuLeu 223

seq_name: pir2:S20590

seq_documentation_block:

exo-alpha-sialidase (EC 3.2.1.18) - Actinomyces viscosus

C:Species: Actinomyces viscosus

C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 22-Oct-1999

C:Accession: S20590

R:Henningsen, M.; Roegentin, P.; Schauer, R.

Biol. Chem. Hoppe-Seyler 372, 1065-1072, 1991

A:Title: Cloning, sequencing and expression of the sialidase gene from Actinomyces viscosus

A:Reference number: S20590; MUID:92162190

A:Accession: S20590

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-913 <HEI>

A:Cross-references: EMBL:X62276; NID:g39254; PIDN:CAA44166.1; PID:g39255

C:Keywords: glycosidase; hydrolase

alignment_scores:

Quality: 122.50 Length: 182

Ratio: 1.376 Gaps: 7

Percent Similarity: 48.901 Percent Identity: 28.571

alignment_block:

US-09-462-480-3 x S20590 ..

Align seg 1/1 to: S20590 from: 1 to: 913

36 CGGACCGCGCGCGCAACCCAGCGAGGAGCGCGAGATGGGCC 85

680 ArgHisArgArgArgSerArgProArgArgProArgArgAlaLeuSerPr 696

86 TGCTCGGCACACAG.....T 99

696 oArgArgHisArgHisHisProArgProSerArgAlaLeuArgProS 713

100 CGGCTGCGAACCATCCGCTGGCTGGATCAGG..... 134

713 erArgAlaGlyProGlyAlaGlyAlaHisAspArgSerGluHisGlyAla 729

135CCCCAGCGCGCGCGCGCTGC 157

730 HisThrGlySerCysAlaGlnSerAlaProGluInThrAspGlyProTh 746

158 TCGCGCGC.....GAGTCGCTACCTGGCGCAGGTGGTGGTGGAC 198
 746 rAlaAlaProAlaProGluThrSerSerAlaProAlaAlaGluProThr 762
 199 CGCAGCGCGCTGATCTCTCAGCTGATCGAAAGCCGGTTGCCCTCGGT 248
 763 GlnAlaProThrValAlaProSerValGluProThrGlnAlaProGlyAl 779
 249 GATCGCGCGCGCTGTTGCCGGATCTCGGTGACGGTGGCGCC...GCTC 295
 779 aglnProSerSerAlaProLysProGlyAlaThrGlyArgAlaProSerV 796
 296 CGGTGGGTCCGGGAGCGATGGCCAGGGTTCGCAATCCGGCGGCTCCACC 345
 796 alValAsnProLysAlaThrGlyAlaAlaThrGluProGlyThrProSer 812
 346 ACCCGCGGTCTGGTGGCGCG.....GCACCGCTCGCGCAGA 383
 813 SerSerAlaSerProAlaProSerArgAsnAlaAlaProThrProLysPr 829
 384 GCGTGAAGAAGACGACGAGGAGCTGGGACGAGGAGGAGGAGCTGGTGAG 433
 829 oglyMetGluProAspGluIleAspArgProSerAspGlyThrMetAlaG 846
 434 CTCCTCGTATGACACAGACTTCCCGGCCACCCGGCGCGCGGAAGA 477
 846 lnPro.....ThrGlyAlaProAlaArgValProArg 857

seq_name: pir2:S51939

seq_documentation_block:

chitinase (EC 3.2.1.14) precursor - beet

C:Species: Beta vulgaris (beet)

C:Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 10-Jul-1998

C:Accession: S51939; S72315; S45025

R:Berglund, L.; Brunstedt, J.; Nielsen, K.K.; Chen, Z.; Mikkelsen, J.D.; Marcker, K.A

Plant Mol. Biol. 27, 211-216, 1995

A:Title: A proline-rich chitinase from Beta vulgaris.

A:Reference number: S51939; MUID:95170004

A:Accession: S51939

A:Molecule type: DNA

A:Residues: 1-439 <BER>

A:Cross-references: EMBL:X79301; NID:g488730; PID:g488731

A:Note: the authors translated the codon TGC for residue 416 as Gly

A:Accession: S72315

A:Molecule type: mRNA

A:Residues: 191-397 <BER2>

C:Genetics:

A:Introns: 248/1; 300/2

C:Superfamily: plant chitinase homology

C:Keywords: glycosidase; hydrolase

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-439/Product: chitinase #status predicted <MAT>

F:183-423/Domain: plant chitinase homology <PCH>

alignment_scores:

Quality: 121.00 Length: 130

Ratio: 2.469 Gaps: 6

Percent Similarity: 37.692 Percent Identity: 33.077

alignment_block:

US-09-462-480-3/rev x S51939 ..

Align seg 1/1 to: S51939 from: 1 to: 439

415 CGTCCAGTCGTCCTCGTCGCTTCTTCACGCTCTCGCGAGCGGTGCC 366

47 ArgProSerArgPro.....ThrProArgProProTh 58

365 GCGCGACACACCGCGGCTGGTGGAGCGCGGATTCGGAACCTGGCC 316

```

58 rProArgProProPro.....ProArgProProThrProArgp 71
315 CATCGTCCCGGACCCACCGAGGCGGCGACCCGCTCACCGACGATCCG 266
   ||| ||||| ||| ||| ||||| |||
71 roProProProProProProProProProProProProArg 87
265 CAACAGCGCGCGGCAATCCGAGGGGGCAACCGGCTTTTCGATCAGCTGA 216
   ||||| ||| ::|||
88 ProProProProArgProProProThrProProArgPro..... 98
215 GACATCAGCGCGTGGGTCACGAC....CCACTCGGGCCAGGTACGG 170
   ||| ||| ||||| |||
99 .....ProProProProProProProProProProProProArgProp 112
169 ACTCGCGCGCAGCAGCGCGCGCGCGCTGGGCGCTGATCCACGACGCC 120
   ::| ||| ||| ||| ::| ||| |||||
112 roThrProArgProProProProProProProProProProProPro 128
119 AGCGGAGTGTTCGACAGCGGACTGGTCGCGAGCAGCCCATCTGCGCGGC 70
   ::| ||| |||||
129 Thr.....ProArgPro..... 132
59 TTCCTCGTGGGTGGCGCGCGCGGTGCGCGGCCACC 30
   ||||| ||||| |||||
133 .....ProProSerProProThr 139

```

THIS PAGE BLANK (uspto)

OM of: US-09-462-480-3 to: SwissProt_40:* out_format : pfs
Date: Jul 22, 2002 1:41 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+n2p_model -DEV=xlh
-O=/cgn2.1/uspro.spool/US09462480/runat-18072002_164420_19600/app_query.fasta_1.2850
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-XGAPOP=4.000 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRLX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM_ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480 -CGN1_1_29 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-462-480-3
Query length: 481
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 49.760000

score_list:	Seq	Strd	Orig	ZScore	Escore	Len
SwissProt_40:WASP_MOUSE	-	145.50	160.09	0.0638	503	1
SwissProt_40:SPPO_HUMAN	-	131.00	142.68	0.4232	707	1
SwissProt_40:GPI_CHRE	-	130.50	143.78	0.4682	555	1
SwissProt_40:ELS_MOUSE	-	127.00	137.19	0.7031	860	1
SwissProt_40:SES_RAT	-	125.50	135.90	0.8644	825	1
SwissProt_40:DIAL_HUMAN	-	120.50	127.92	1.59	1248	1
SwissProt_40:EXTN_TOBAC	-	119.00	131.03	2.15	620	1
SwissProt_40:FMN_CHICK	-	118.00	125.50	2.23	1213	1
SwissProt_40:YK82_MYCTU	-	117.50	128.46	2.57	721	1
SwissProt_40:EXLP_TOBAC	-	117.00	131.46	2.96	426	1
SwissProt_40:IE18_PRVKA	-	117.00	123.28	2.49	1446	1
SwissProt_40:TPM5_DROME	-	115.50	128.77	3.31	504	1
SwissProt_40:AFSK_STRGR	-	115.50	125.62	3.33	807	1
SwissProt_40:VNUA_PRVKA	-	115.50	120.50	2.97	1733	1
SwissProt_40:APG_ARATH	-	115.00	127.86	3.75	534	1
SwissProt_40:T2D3_HUMAN	-	115.00	123.13	3.39	1083	1
SwissProt_40:HMID_DROAN	-	114.50	126.49	3.94	606	1
SwissProt_40:SEPA_EMENI	-	114.50	119.24	3.38	1790	1
SwissProt_40:MYCN_WARMO	-	114.00	127.81	4.38	460	1
SwissProt_40:IE18_PRVIF	-	114.00	120.08	3.72	1461	1
SwissProt_40:Y735_STRFR	-	113.50	129.16	4.87	348	1
SwissProt_40:KLF2_HUMAN	-	113.50	129.03	4.85	355	1
SwissProt_40:CA21_RANCA	-	113.50	120.06	4.02	1355	1
SwissProt_40:TEGU_HSV11	-	113.50	114.38	3.56	3164	1
SwissProt_40:HK61_RAT	-	112.50	127.80	5.53	365	1
SwissProt_40:WASP_MOUSE	-	112.50	125.43	5.26	520	1
SwissProt_40:CAPU_DROME	-	112.50	120.67	4.75	1059	1
SwissProt_40:DRPL_RAT	-	112.50	119.92	4.68	1183	1
SwissProt_40:CA21_BOVIN	-	112.00	118.97	4.59	1364	1
SwissProt_40:D4DR_HUMAN	-	112.00	125.63	5.71	467	1
SwissProt_40:SYNL_MOUSE	-	112.00	123.21	5.42	670	1
SwissProt_40:SYNL_HUMAN	-	112.00	122.87	5.38	705	1
SwissProt_40:NKCL_MOUSE	-	112.00	119.28	4.99	1205	1
SwissProt_40:ICP3_HSVIF	-	111.50	128.95	6.62	263	1
SwissProt_40:NO20_MEDTR	-	111.50	128.82	6.60	268	1
SwissProt_40:EBN1_EBV	-	111.50	122.98	5.83	641	1
SwissProt_40:SYNL_BOVIN	-	111.50	122.34	5.76	706	1
SwissProt_40:CSP_PLASI	-	111.00	125.86	6.70	386	1
SwissProt_40:Y70K_TYMCV	-	111.00	122.60	6.26	628	1
SwissProt_40:CCAB_RABIT	-	111.00	113.80	5.20	2339	1

SwissProt_40:ICP3_HSV11	-	110.50	128.30	7.63	248	1	P36313 herpes simplex virus 1
SwissProt_40:YEN1_SCHPO	-	110.50	123.14	6.84	536	1	O13695 schizosaccharomyces pombe
SwissProt_40:NCR2_HUMAN	-	110.50	112.78	5.50	2517	1	Q9Y618 h nuclear receptor
SwissProt_40:CSP_PLAKU	-	110.00	125.45	7.77	351	1	P04922 plasmodium knowlesi
SwissProt_40:HK61_MESAU	-	110.00	125.21	7.73	364	1	Q60554 mesocricetus auratus

seq_name: SwissProt_40:WASP_HUMAN

seq_documentation_block:

ID	WASP_HUMAN	STANDARD;	PRT;	503	AA.
AC	043516; OSUNP1; Q15220;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Wiskott-Aldrich syndrome protein interacting protein (WASP interacting protein) (PRPL-2 protein).				
DE	WASP or WIP.				
GN	Homo sapiens (Human).				
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=98070810; PubMed=9405671;				
RA	Ramesh N., Anton I.M., Hartwig J.H., Geha R.S.;				
RT	"WIP, a protein associated with Wiskott-Aldrich syndrome protein, induces actin polymerization and redistribution in lymphoid cells.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 94:14671-14676(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Tonsil;				
RA	Kreideweiss S., Delany-Heiken P., Nordheim A., Ruhlmann A.;				
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 192-503 FROM N.A., AND VARIANT ALA-495.				
RX	MEDLINE=99218549; PubMed=10202051;				
RA	Stewart D.M., Tian L., Nelson D.L.;				
RT	"Mutations that cause the Wiskott-Aldrich syndrome impair the interaction of Wiskott-Aldrich syndrome protein (WASP) with WASP interacting protein.";				
RL	J. Immunol. 162:5019-5024(1999).				
CC	-!- FUNCTION: MAY HAVE DIRECT ACTIVITY ON THE ACTIN CYTOSKELETON.				
CC	-!- INDUCES ACTIN POLYMERIZATION AND REDISTRIBUTION.				
CC	-!- SUBUNIT: BINDS TO WASP, PROFILIN AND ACTIN.				
CC	-!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PERIPHERAL BLOOD MONONUCLEAR CELLS, SPLEEN, PLACENTA, SMALL INTESTIN, COLON, THYMUS. LOWER EXPRESSION IN OVARY, HEART, BRAIN, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS, PROSTATE, TESTIS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AF031588; AAC03767.1; -				
DR	EMBL; AF106062; RAD45972.1; -				
DR	EMBL; X86019; CAA60014.1; -				
DR	MIM; 602357; -				
DR	InterPro; IPR003124; WH2.				
DR	Pfam; PF02205; WH2; 1.				
DR	SMART; SM00246; WH2; 1.				
KW	Actin-binding; Repeat.				
FT	DOMAIN 2 13	POLY-PRO.			
FT	DOMAIN 64 96	GLY-RICH.			
FT	DOMAIN 241 244	POLY-SER.			
FT	DOMAIN 264 433	PRO-RICH.			
FT	REPEAT 352 361	XRSGPXPXP MOTIF 1.			
FT	REPEAT 374 383	XRSGPXPXP MOTIF 2.			
FT	REPEAT 410 419	XRSGPXPXP MOTIF 3.			
FT	SITE 45 48	ACTIN BINDING.			

```
FT VARIANT 495 495 G -> A.
FT CONFLICT 303 309 /FTID=VAR_010295.
FT CONFLICT 360 360 PHRHLR -> SSQAPP (IN REF. 3).
FT CONFLICT 487 503 P -> PV (IN REF. 2).
FT CONFLICT 487 503 GSNRRRGGPPLPIPR -> EYFCQGF (IN REF. 2).
SQ SEQUENCE 503 AA; 51489 MW; 43EB88674DD3BF1A CRC64;

alignment_scores:
  Quality: 145.50 Length: 163
  Ratio: 2.021 Gaps: 10
  Percent Similarity: 44.172 Percent Identity: 33.742

alignment_block:
US-09-462-480-3/rev x WAIP_HUMAN ..
Align seg 1/1 to: WAIP_HUMAN from: 1 to: 503

472 CGGCGCGGGTGGCGGGAGTCTGTGTATACGGAGCTCACAGTCG 423
|||||
315 ArgProGlyProPro.....
422 TCCCTTCGTCCTCCAGTCGT.....CCTC 400
|||||
320 .ProleuProProSerSerSerGlyAsnAspGluThrProArgLeuProG 336
|||||
399 GTCGTCTTCTTCACGCTCCTCGCGAGCGGTGCGCGCGCCGACACGACCG 350
|||||
336 InArgAsnLeu...SerLeuSerSerThrProProLeuProSerPro 351
|||||
349 GCGTGTGGAGCCCGCGGATTGCGAACCTGCGCCATCGCTCCCGGACCC 300
|||||
352 Gly.....ArgSerGlyPro...LeuProProPr 360
|||||
299 ACCGAGCGGGCCACCGTCACCGACGATCCGCGCAACAGCCGCGGCAT 250
|||||
360 oProSerGluArg...ProProProValArgSerProGlyArgS 376
|||||
249 CA.....CCGAGGGGCAACCGGCTTTTCATCAGTGCAGACATCA 209
|||||
376 erGlyProProProProProProValSerArgAsnGlySerThrSer 392
|||||
208 CGCGCGTGGGTCAAGACCCACCTCGGCA.....GTAGGAC 168
|||||
393 ArgAlaLeuProAlaThr...ProGlnLeuProSerArgSerGlyValAsp 408
|||||
167 TCCGCGCGCAGCAGCGCGCGCGCGCTGCGGCTGATCCACCGACG 118
|||||
409 SerProArgSerGlyProArgProProLeuProProAspArgPro..... 423
|||||
117 CGGATGTTGACACAGCGGACTGTGTGCGGAGCAGCGCCCATCTGCGCGGCTT 68
|||||
423 .....
67 CCTCGTGGCTGGTGGCGCGCGCGCGTGGCGGCCACCC 30
|||||
424 .....SerAlaGlyAlaProProProProProSer 434
|||||

seq_name: SwissProt_40:SFPO_HUMAN

seq_documentation_block:
ID SFPO_HUMAN STANDARD; PRT; 707 AA.
AC P23246; P30808;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor, proline- and glutamine-rich (Polypyrimidine tract-
DE binding protein-associated splicing factor) (PTB-associated splicing
DE factor) (PSF) (DNA-binding p52/p100 complex, 100 kDa subunit).
GN SFPO OR PSF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal brain;
RX MEDLINE=93194059; PubMed=8449401;
RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
RT factor.";
RL Genes Dev. 7:393-406(1993).
RN [2]
RP SEQUENCE OF 312-707 FROM N.A.
RC TISSUE=Fetal skeletal muscle;
RX MEDLINE=90091812; PubMed=2480877;
RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
RT "Cloning and characterization of a myoblast cell surface antigen
RT defined by 24.1D5 monoclonal antibody.";
RL Development 105:723-731(1989).
RN [3]
RP SEQUENCE OF 48-68 AND 213-246.
RX MEDLINE=93176127; PubMed=8439294;
RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
RT "Purification and characterization of a DNA-binding heterodimer of 52
RT and 100 kDa from HeLa cells.";
RL Biochem. J. 290:267-272(1993).
CC -!- FUNCTION: ESSENTIAL PRE-MRNA SPLICING FACTOR REQUIRED EARLY IN
CC SPLICEOSOME FORMATION. BINDS TO THE MAMMALIAN POLYPYRIMIDINE
CC TRACTS. FORMS A COMPLEX WITH THE POLYPYRIMIDINE TRACT-BINDING
CC PROTEIN (PTB). SEEMS TO ALSO BIND DNA.
CC -!- SUBUNIT: HETEROTETRAMER OF TWO 52 KDA AND TWO 100 KDA SUBUNITS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -!- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL
CC SURFACE ANTIGEN 24.1D5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN
CC ECTOKINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X70944; CAA50283.1; -
DR EMBL; X16850; CAA34747.1; -
DR PIR; A43557; A43557.
DR PIR; S29770; S29770.
DR HSP; P19339; 1SXL.
DR MIM; 605199; -
DR InterPro; IPR000504; RRM.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
DR PROSITE; PS00303; RRM_RNP_1; 1.
KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
KW Alternative splicing.
FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
FT DOMAIN 9 27 3 X 3 AA REPEATS OF R-G-G.
FT REPEAT 9 11 1.
FT REPEAT 19 21 2.
FT REPEAT 25 27 3.
FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
FT DOMAIN 10 15 POLY-GLY.
FT DOMAIN 20 27 POLY-GLY.
FT DOMAIN 56 65 POLY-PRO.
FT DOMAIN 67 71 POLY-GLN.
FT DOMAIN 95 98 POLY-GLN.
FT DOMAIN 99 103 POLY-PRO.
FT DOMAIN 184 188 POLY-PRO.
FT DOMAIN 571 574 POLY-ARG.
```

219 CTGAGACATCAGCGGGTGGGTCAACGGA

TT8PROPSEIPROPSEIPROPSEI..SEIPIOSE 130

RA Ferris P.J., Woessner J.P., Waffenschmidt S.,


```
alignment_scores:
  Quality: 125.50      Length: 198
  Ratio: 1.294         Gaps: 10
  Percent Similarity: 48.990  Percent Identity: 31.818

alignment_block:
US-09-462-480-3 x 5E5_RAT ..

Align seg 1/1 to: 5E5_RAT from: 1 to: 825

24 CAGCCAGGTGGG.....CGGCACCGCGCGCGGCAACCCACGACGACGAGG 67
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
478 GlnAlaGlyGlyGlyProArgGlyArgAlaArgGlyProArgGlyGlnAla 494
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 AAGCCGCGCAGATGGGCTCTCGGCACACAGTCCGCTGTCGAACCATCCG 117
   : ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
494 aArgArgGlyHisGlyProGlnArgArgGly.....Pro.Pro 507

118 CTGGCTGGTGGATCAGGCCCGCAGCGCGCGCGGCGCTGCTG..... 159
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
508 GlnAlaGlyGluGluGlyProGlyAspAlaThrLeuValLeuGlyLeuG 524

160 .....CGCGCGAGTCGCTACCTGGCG 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
524 yThrThrSerGlyGluGlnArgAlaAspGlnSerGlnThrLeuProAlaL 541
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
182 CAGTGGTGGTGCACCG.CAGCCGCTGATGTCTCAGCTGATCGAATAA 230
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
541 euAlaGlyAlaProThrAlaHisAlaHisAlaValProGlyProGly 557

231 GCCGGTGGCCCGCTGGTATGCGGC.....GGCTGTGGCGGATCGT 274
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
558 AlaAlaAlaThrLeuGlyGlyArgGlyArgGlySerTrpArg..... 572

275 CGGTGACGGTGGCGCGCTCGGTGGTCCGGAGCGATGGCGACGGT 324
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
573 .....GlyGlyArgArgGlyGlyGlyAlaGlyAlaSerGlyGlyAla 587

325 TCGCAATCCG.....CGGCTCCACGACCGCGGTGTGTGCGGCC 365
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
587 rgGlyGlyArgGlyArgGlyArgGlyArgGlySerGlyLeuSer 603

366 GGCACC.....GCTCGCGCAGGAGCGGTGAAG 391
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
604 GlyThrArgGluAspAlaGlySerProSerAlaArgArgGlyGluGln 620

392 AAGACGA.....C 399
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
620 gArgGlyHisGlyProProAlaAlaGlyAlaAlaGlnValSerThrA 637

400 GAGCAGCACTGGGACGAGGAGGAGGAGTGTGAGCTCCGCTAATGAC... 446
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
637 rgGlyArgArgAlaArgGlyGlnArgThrGlyGluGluAlaGlnAspGly 653

447 .....AACAGACTTCCCGCCACCGCGGCG 470
   : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
654 LeuLeuProArgGlyArgAspArgLeuProLeuArgProGly 667

seq_name: SwissProt_40: DIAL_HUMAN

seq_documentation_block:
ID DIAL_HUMAN STANDARD; PRT: 1248 AA.
AC Q06010; Q9UC76;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Diaphanous protein homolog 1 (Diaphanous-related formin 1) (DRF1).
GN DIAPH1 OR DIAPI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```


470	GCCCGGGTGGCCGGGAAGTGTTGTCATTACGGGAGCTCACCGACTGCCT	421
:::	:::::::::::::::::::: :::	
318	AlaLysAlaArgAlaGluGluLeuAlaAlaLeuGlyGluGluAlaGlyAl	334
420	CYTCTCGTCCCAGTCTGCTCTGCTGCTCTTCAACGCTCCTCGCGCCAGCG	371
:::	::: :::	
334	aGluAlaGlyGluGlyGlyValaProAlaGluGlyAlaAalaProGlyGluP	351
370	GTGCCGGCGGCACACACCGCGCTGGTGAGCGCCGATTCGGAACACC	321
:::	::: :::	
351	roGlyAlaIatHrGluProGly...ValGluAlaProAlaGluPro	366
320TGGCCCCATCGC	310
:::	:::	
367	GluAtgIleProThrProProProProPheGluTyrSerIleAs	383
309	TCCGGACCCACGGGAGCG..	291
383	pLeuProProGluGlyAlaGluValProTyrValArgAsnAlaGluProG	400
290GCGCCACCGCTCACCGAGATCCGGCAACAGCCGCGCATC	249
400	IyAspPheAlaProAlaGluAlaAlaProAlaGluGlyAlaProPro	416
248	ACGAGGGGGCAACCGGCTTTTCGATCAGCTGAGACATCAGCGCGCTGCG	199
:::	::: :::	
417	AlaGluGlyAlaProAlaAlaGluGlyAlaAlaProAlaGluGlyAlaPr	433
198	GGTCACAGAC.....CCACTGGCGCAGTAGCGACTCCGCGCGCAGCA	155
:::	:::	
433	oAlaAlaGluGlyAlaProProAlaGluGlyAlaProAla	446
154	GCGCGCGCGCGCGCTGGGS.....CCTGATCCACGACGACGC	117
447	.ProAlaProAlaGluGlyGluAlaAlaProProAlaProAlaGlu	462
116	GGATGTTTCAGACGAGCATGTGTCGCGAGCGGCC.....AT	79
463	Gly.....AspAlaAlaAlaProProProProAlaGluGly	477
78	CTGCGCGGCTTCTCGTGGTGGTGGTGGCGCGCGGTCGCGCC	33
:::	::: :::	
477	uAlaAlaProAlaProAlaGluGlyAlaProAlaProAlaPro	492
seq_name:	SwissProt_40:AFSK_STRGR	
seq_documentation_block:		
ID	AFSK_STRGR	STANDARD;
AC	P54742;	PRT; 807 AA.
DT	01-OCT-1996	(Rel. 34, Created)
DT	01-OCT-1996	(Rel. 34, Last sequence update)
DT	30-MAY-2000	(Rel. 39, Last annotation update)
GN	Serine/threonine protein kinase afsk	(EC 2.7.1.-).
OS	Streptomyces griseus.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OX	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces	
NCBI_TaxID=1911;		
[1]	SEQUENCE FROM N.A.	
RX	MEDLINE=96186909; PubMed=8635757;	
RA	Ueda K., Umeyama T., Beppu T., Horinouchi S.;	
RT	"The aerial mycelium-defective phenotype of Streptomyces griseus	
RT	resulting from A-factor deficiency is suppressed by a Ser/Thr Kinase	
RL	of S. coelicolor A3(2).";	
Gene	169:91-95(1996).	
-!	FUNCTION: INVOLVED IN THE REGULATION OF SECONDARY METABOLISM BY	
CC	PHOSPHORYLATING, ON BOTH SER AND THR, THE AFSR GLOBAL REGULATORY	
CC	PROTEIN INVOLVED IN THE CONTROL OF SECONDARY METABOLISM (BY	
CC	SIMILARITY).	
-!-	PTM: AUTOPHOSPHORYLATED AT THR AND SER RESIDUES (BY SIMILARITY).	
CC	-!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.	
CC	-----	
CC	CC	

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC	EMBL; D45246; BAA08203.1; -								
DR	InterPro; IPR000719; Euk_pkinase.								
DR	InterPro; IPR002290; Ser_thr_pkinase.								
DR	Pfam; PF00069; pkinase; 1.								
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.								
DR	PROSITE; PS00108; PROTEIN_KINASE_ST; 1.								
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.								
KW	Transferase; Serine/threonine-protein kinase; ATP-binding; Phosphorylation.								
KW									
FT	DOMAIN	16	272						PROTEIN KINASE.
FT	NP_BIND	22	30						ATP (BY SIMILARITY).
FT	BINDING	44	44						ATP (BY SIMILARITY).
FT	ACT_SITE	138	138						BY SIMILARITY.
SQ	SEQUENCE	807	AA; 85231	MW; 666274219155D091	CRC64;				

```
alignment_scores:
  Quality: 115.50      Length: 163
  Ratio: 1.44         Gaps: 10
  Percent Similarity: 49.080  Percent Identity: 33.742
```

```
alignment_block:
US-09-462-480-3/rev x AFSK_STRGR
```

Align seq 1/1 to: AFSK_STRGR from: 1 to: 807

439 CGGGAGCTCACCAGTCGTCTTTCGTCACCATCGTCGTCGTCGCTCTCT 390
||||: :: ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 ArgArgThrAlaArgArgProArgPro...ArgProArgLeu... 314
389 TCACGCTCTCGCGAGCGTGCCGGCGGCACCAAGACCCGGCTGGTGA 340
||||: ||| ||| ||| |||
315 ...ArgAlaAlaPro.GlnGlyPro 321
339 GCGCGCGGATTGGAACCTGGCCATCTCCGAGCACCCAGCGAGCGG 290
:::||||| ||| |||||:::|||||:::|||||:::|||||
322 GlyAlaGlyHisArg...LeuAlaGlnArgGlyArg.ProAlaPheAlaL 337
289 GCGCACCGCTCACCGACGATCCGGCA.....ACAGCGCGC 255
||| ||| ||| |||
337 euProAlaValLeuAlaLeuAlaAlaValArgValArgThrAlaAla 353
254 GGCAATCAGCGGGGGCAACCGGCTTTTCGATCAGCTGAGATCATCAGCGG 205
||| :||: ||| ||| :||: |||
354 GlyProSerAlaAlaProAspGlyGlyProVal.....GlnLeuProGl 368
204 CGTGCGGGTCAACGACCACTCGGCCAGGTAGCACTCCGGCGCGCAGCA 155
||||: ||| ||| :||: |||||
368 yAlaLysVal.....ProIleclYProGlyArgAlaGlyGluGlyA 383
154 GGCGCGGGCGCGG.....CTGGGGCTGTAT 129
||| ||| ||| |||||
383 rgGlyAlaAlaAlaAlaProArgArgAspArgLeuGlyProAla 399
128 CCACCAAGCAGCGGATGGTTTCAGACGCGACTGGTGCCGAGCAGGCCCAT 79
||| ||| ||| :||: ||||| ||| ||| :||: |||
400 AlaArgSerGluArgPheLeuGlyGlyHisGlyProHisArgThrVa 416
78 TCGCGCGGCTTCCTCCTCGGCTGGTTCGCGCGCGGTGCGCGCCACCT 29
: |||: ||| ||| |||||
416 lProAlaSerThrLeuArgProGlyThrProSerProAlaProAspArgT 433
28 GG.....CTGACAACGACGCTC 12
|| ||| ||| |||||
433 rPargProTrpArgPheArqMetSerAsnAspVal 444

158 TCGCGCGGAG.....TCGTACT 177
:: |||||
923 alProAlaGlyArgAlaGlyGluProLeuValLeuLeuAlaValPro 939
178 GCGCAGTGGTGTGTCACCGCAGCGCGTGTCTCAGCTGATCGA 227
|||||
940 GlyAlaGly.....ProGlyArgAlaAlaLeuLeu 951
228 AAAGCGGTTCGCCCTCGGTGATCGCGCGCTGTTCGCGATCGTCGG 277
|||||
951 uAlaProLeuGlyArgTrpValArg...AlaGlyGlyGlyAlaGlyV 967
278 TCACGGTGGCGCGCTCCGTGGTCCGGAGCG..... 312
|||||
967 alAlaGlyGlyAlaGlyGluAlaGlyLeuGlyAlaGlyAlaGly 983
313 ATGGCCAGGTTTCGATCCGGCGGTCCAC.CAGCCCGGTCTGTGTCG 361
|||
984 AlaGlyAlaGlyLeuGlyAlaGlyGlyAlaGlyProGlyAlaGly 1000
362 CGCGGCGACCGCTCGCGCAGGCGGTGAAGAGCAGCAGGAGCTGG 411
:|||||
1000 uAlaGly.....GlyGlyAlaArgArgArgArgArgArg.Trp 1013
412 GACGAAGAG 420
|||||
1014 AspAspGlu 1016

seq_name: SwissProt_40:APG_ARATH

seq_documentation_block:

ID APG_ARATH STANDARD; PRT; 534 AA.
AC P40602;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Anter-specific proline-rich protein APG precursor.
GN APG.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94004980; PubMed=8401599;
RA Roberts M.R., Foster G.D., Blundell R.P., Robinson S.W., Kumar A.,
RA Draper J., Scott R.;
RT "gametophytic and sporophytic expression of an anther-specific
Arabidopsis thaliana gene.";
RL Plant J. 3:111-120(1993).
CC -!- TISSUE SPECIFICITY: FOUND IN SPOROPHYTIC AND GAMETOPHYTIC CELL
TYPES IN THE ANTHUR, ONLY IN MALE FERTILE PLANTS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS, DURING
MICROSPORE DEVELOPMENT. HIGHER EXPRESSION IS FOUND DURING
MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN
MATURATION.
CC -!- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X60377; CAA42925.1;
DR PIR; S21961; S21961.
DR InterPro; IPR001087; Lipase_GDSL.
DR Pfam; PF00657; Lipase_GDSL; 1.
DR PROSITE; PS01098; LIPASE_GDSL_SER; 1.
KW Signal.

FT SIGNAL 1 35
FT CHAIN 36 534
FT ACT_SITE 211 211
FT ACT_SITE 511 511
SQ SEQUENCE 534 AA; 57967 MW; 744CAD3B08CC482E CRC64;

POTENTIAL.
ANTER-SPECIFIC PROLINE-RICH PROTEIN APG.
BY SIMILARITY.
POTENTIAL.

alignment_scores:
Quality: 115.00 Length: 154
Ratio: 1.691 Gaps: 9
Percent Similarity: 44.156 Percent Identity: 29.221

alignment_block:

US-09-462-480-3/rev x APG_ARATH

Align seg 1/1 to: APG_ARATH from: 1 to: 534

463 TGGCGGGAAGTCTGTTCATTACGGAGCTCACCAGTCTCCTCTTCG 414
|||||
43 TrpProTrpProLeuTrpProArgProTrpProGlnProTrpProMetAs 59
413 TCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 364
:|||||
59 nProProThrProAsp.....ProSerProLysProValAlaPro 73
363 CGCGACCAAGACCCCGGCTGTGTGGAGCCCGCGGATTCGGAACCTGGCCCA 314
|||||
73 roGlyProSerPro.....LysProValAlaProProGlyPro 85
313 TCG...CTCCCGACCCACCGGAGCGGCCAC..... 284
|||
86 SerProCysProSerProProLysProGlnProLysProProAl 102
283 .CCGTCACCGAGATCCGGCAACAGCCCGCGCATCACCGAGGGGCAAC 235
|||||
102 aProSerProSerProCysProSerProProLysProGlnProLysP 119
234 CGGCTTTTCGATCAGCTGAGACATCAGCGCGTGTGGTCAACGCCAC 185
|||
119 roVal.....ProProProAlaCysProProThr...Prop 130
184 CTGCGCCAGGTAGGACTCCCGCGCGCAGCAGCGCGCGCGCGCTGGG 135
|||
130 roLysPro.....GlnProLysProAlaProProProAlaProLys 143
134 CCTGATCCACGACGCGGATGTTCGACAGCGGACTGGTGGCGAGCAG 85
|||
144 ProAlaProProAla.....ProLys 151
84 GCCCATCTGCGCGCTTCC.....TCGTCGGCTGGGTGGCGCGCG 44
:|||||
151 sProValProCysProSerProProLysProProAlaProThrProLysP 168
43 CGGTGGCGGCC 33
|||||
168 roValProPro 171

OM of: US-09-462-480-3 to: SPTREMBL_19:* out_format : pfs
Date: Jul 22, 2002 1:40 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+n2p_model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09462480/runat-18072002_164419_19544/app_query.fasta_1.2850
-DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOB=12.000
-GAPOPT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-FGAPOPT=4.500 -FGAPOPT=0.050 -XGAPOB=10.000 -XGAPEXT=0.500
-FGAPOB=6.000 -FGAPEXT=7.000 -XGAPOB=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -NATRLX=blosum62
-TRANS=human40.cgi -LIST=45 -DLOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480 -CGN1_1_133 -ICPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-3
Query length: 481
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 143.350000

score_list:

Sequence	Strd	Orig	zScore	EScore	Len	Documentation
sp_bacteriap:O69738	+	713.00	955.62	1.9e-45	368	O69738 mycobacterium tuberculosis
sp_bacteriap:O33085	+	198.00	266.81	5.4e-07	302	O33085 mycobacterium leprae. hy
sp_bacteriap:Q9E9Z2	-	143.50	197.95	0.0044	251	Q9E9Z2 paracoccus pantotrophus.
sp_human:Q9BU37	-	138.00	184.97	0.0164	358	Q9BU37 homo sapiens (human). si
sp_bacteriap:Q9DXH2	-	134.00	173.38	0.0341	763	Q9DXH2 mycobacterium tuberculosis
sp_bacteriap:Q59164	-	133.00	170.56	0.0408	913	Q59164 actinomycetes viscosus. si
sp_plant:Q9FPQ6	-	130.50	171.31	0.0611	555	Q9FPQ6 chlamydomonas reinhardt
sp_mammal:Q9N1P0	-	128.50	165.43	0.0880	818	Q9N1P0 bos taurus (bovine). sub
sp_invertebrate:Q9NHW4	+	128.50	157.11	0.0931	2249	Q9NHW4 neophila clavipes (orb
sp_virus:IP89459	-	128.50	154.41	0.0948	3122	IP89459 herpes simplex virus (b
sp_invertebrate:O44358	+	128.00	164.25	0.0963	871	O44358 neophila clavipes (orb
sp_rident:Q9WUE8	-	127.50	156.39	0.1101	2087	Q9WUE8 rattus norvegicus (rat)
sp_rident:Q9WU13	-	127.50	156.11	0.1103	2158	Q9WU13 rattus norvegicus (rat)
sp_rident:Q9WV48	-	127.50	156.08	0.1103	2167	Q9WV48 rattus norvegicus (rat)
sp_rident:Q9ES29	+	127.00	163.50	0.1139	810	Q9ES29 mus musculus (mouse). el
sp_plant:Q9SX31	-	125.50	162.60	0.1463	708	Q9SX31 arabidopsis thaliana (mc
sp_bacteriap:Q9F2N5	-	125.00	160.06	0.1614	889	Q9F2N5 streptomyces coelicolor.
sp_invertebrate:Q20739	-	124.50	168.22	0.1658	304	Q20739 caenorhabditis elegans
sp_invertebrate:Q9VZC2	-	124.50	165.06	0.1693	446	Q9VZC2 drosophila melanogaste
sp_bacteriap:Q33268	-	124.50	159.45	0.1758	882	Q33268 mycobacterium tuberculosis
sp_bacteriap:O69995	-	124.00	160.18	0.1898	744	Q69995 streptomyces coelicolor.
sp_bacteriap:O06556	-	123.00	168.61	0.2111	227	O06556 mycobacterium tuberculosis
sp_virus:Q9IPQ8	+	123.00	160.78	0.2225	588	Q9IPQ8 cynomolgus epstein-barr
sp_mammal:Q9N1P1	-	123.00	160.20	0.2234	631	Q9N1P1 bos taurus (bovine). sub
sp_bacteriap:Q59164	+	122.50	156.49	0.2485	913	Q59164 actinomycetes viscosus. si
sp_invertebrate:Q9VQ94	-	122.50	154.29	0.2522	1192	Q9VQ94 drosophila melanogast
sp_invertebrate:Q96016	-	122.50	154.29	0.2522	1193	Q96016 drosophila melanogast
sp_human:Q96508	-	122.00	153.68	0.2747	1184	Q96508 homo sapiens (human). s
sp_invertebrate:Q9NHW2	+	122.00	149.86	0.2818	1884	Q9NHW2 neophila madagascari
sp_rident:Q88493	-	122.00	147.03	0.2873	2657	Q88493 mus musculus (mouse). si
sp_bacteriap:Q68872	-	121.50	159.44	0.2867	542	Q68872 myxococcus xanthus. hypc
sp_rident:Q9BK31	-	121.50	158.38	0.2887	616	Q9BK31 mus musculus (mouse). si
sp_human:Q9F2P0	-	121.50	153.22	0.2989	1154	Q9F2P0 homo sapiens (human). k
sp_plant:O42421	-	121.00	160.50	0.3088	439	Q42421 beta vulgaris (sugar bee
sp_rident:Q62775	-	121.00	160.28	0.3093	451	Q62775 rattus norvegicus (rat).
sp_virus:Q69270	-	121.00	159.66	0.3099	469	Q69270 equine herpesvirus 1. ir
sp_rident:Q920G8	-	121.00	159.68	0.3105	485	Q920G8 rattus norvegicus (rat).
sp_plant:Q9FU27	-	121.00	157.05	0.3161	688	Q9FU27 zea mays (maize). scared
sp_rident:Q9CTW8	-	120.50	163.80	0.3277	271	Q9CTW8 mus musculus (mouse). 49
sp_invertebrate:Q9BIV0	+	120.50	161.72	0.3323	349	Q9BIV0 argiope aurantia. majfd

seq_name: sp_bacteriap:O69738
seq_documentation_block:
ID O69738 PRELIMINARY; PRT; 368 AA.
AC O69738;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PPE-FAMILY PROTEIN
GN RV3873 OR MTV027.08.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RAC Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares R., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
DR EMBL; AL022120; CAA17965.1; -.
DR Tuberculist; RV3873; -.
DR InterPro; IPR000030; PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 368 AA; 37330 MW; D78F44095F658CA2 CRC64;
alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057
Percent Similarity: 98.601 Percent Identity: 96.503
alignment_block:
US-09-462-480-3 x O69738 ..
Align seg 1/1 to: O69738 from: 1 to: 368
1 CTGCAGAGTGTGCTGTGTTTCAGCAGTGGCGGCGACCGCGCGG 50
|||||
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCGACGACGAGGAGCGCGGCGGCTGCTCGGCACCA 100
|||||
242 YASNProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThr 259
101 CGCTGTGCAACCATCCGCTGCTGGTGGATCAGGCCCGCGGCG 150
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAla 275
151 GGCTGTGTCGCGCGGAGTGCCTACCTACCTGGCGCGAGTGGT 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySer 292
201 CAGCCCGCTGATGCTCTCAGCTGATCGAGAAAGCGGTTGCC 250
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaPro 309


```

366 CGCGG.....CGACACAGCCGGCTGGTGGAGC...CCCGCGATT 329
   ::::: ||||| ||| ||| ||| |||||:::
113 laSerArgLysAsnArgProLysProPrrpArgSerArgArgVal 129
   ::::: ||||| ||| ||| ||| |||||:::
328 GCGAACCTCGGC.....CCATCGCTCCCGACCCACCGGA 294
   ::::: ||||| ||| ||| ||| |||||:::
130 SerLysProSerArgArgSerArgProSerSerLeuProArgProH1 146
   ::::: ||||| ||| ||| ||| |||||:::
293 GCGGCGCCACCGCTACCGACGATCCGGCAACACGCGCGCGCATCACCGA 244
   ::::: ||||| ||| ||| ||| |||||:::
146 saArgThrProArgArgHisArgMetArgArgProGlySer.... 161
   ::::: ||||| ||| ||| ||| |||||:::
243 GGGGGCAACCGGCTTTTCGATCAGCTGACATCAGCGGGCTGGGTCA 194
   ::::: ||||| ||| ||| ||| |||||:::
162 .....ArgLeuProSerLeuProSerAlaArgArgPro 172
   ::::: ||||| ||| ||| ||| |||||:::
193 ACACACACCTCGCGCAGGTAGGACTCCGCGCGCAGCAGCGCGCGCC 144
   ::::: ||||| ||| ||| ||| |||||:::
173 ThrProArgLeuLeuAsnProProArgProThrProArgLysProArgPr 189
   ::::: ||||| ||| ||| ||| |||||:::
143 CGCGTGGGGCTGATCCACGACGCGGATGTTTCGACAGCGGACTGGT 94
   ::::: ||||| ||| ||| ||| |||||:::
189 oArg..AsnProLysPro.....SerGlyGlnAr 198
   ::::: ||||| ||| ||| ||| |||||:::
93 GCGGACGAGCCCATC.....TCGCGGGCTTCCTCGTGGGCTG 56
   ::::: ||||| ||| ||| ||| |||||:::
198 gProLysArgProLeuHisProLysProCysArgArgProSerProSerP 215
   ::::: ||||| ||| ||| ||| |||||:::
55 GGTTCGCGCGCGCGTGGCGGCC 33
   ::::: ||||| ||| ||| ||| |||||:::
215 roProSerProLeuProPro 222
   ::::: ||||| ||| ||| ||| |||||:::

```

seq_name: sp_human:Q9BU37

```

seq_documentation_block:
ID Q9BU37 PRELIMINARY; PRT: 358 AA.
AC Q9BU37;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE SIMILAR TO WISKOT-ALDRICH SYNDROME PROTEIN INTERACTING PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA.;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002914; AAH02914.1; -.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF02205; WH2; 1.
DR PRINTS; PR01217; PRICHEXTNSN.
DR SMART; SM00246; WH2; 1.
SQ SEQUENCE 358 AA; 36464 MW; D008B60E60EE94EA.CRC64;

```

alignment_scores:
 Quality: 138.00 Length: 155
 Ratio: 2.226 Gaps: 6
 Percent Similarity: 40.000 Percent Identity: 31.613

alignment_block:

US-09-462-480-3/rev x Q9BU37 ..
 Align seg 1/1 to: Q9BU37 from: 1 to: 358

```

412 CCAGTCGCTCGTCTCTCTTCAGCTCCTCGCGAGCGGTGCGGC 363
   ||| ::::: ||||| ||| ||| ||| |||||:::
161 ProProGluProGlnArg.....AsnArgMetProProPr 172

```

```

362 GCGACACAGCCGGCTGGTGGAGCCCGCGATTGCGAACCCCTGGCCCAT 313
   ||||| ||| ||| ||| |||||:::
172 oArgProAspValGlySerLysProAspSerIleProProValProS 189
   ::::: ||||| ||| ||| ||| |||||:::
312 CGTCCCGGACCCA..... 299
   || ||| |||
189 erThrProArgProIleGlnSerSerLeuHisAsnArgGlySerProPro 205
   ::::: ||||| ||| ||| ||| |||||:::
298 .....CCGGAGCGCGCCACCGCTACCGACGATCCGCGCAACA 261
   ||||| ||| ||| ||| |||||:::
206 ValProGlyProArgGlnProSerProGlyProThrProProPr 222
   ::::: ||||| ||| ||| ||| |||||:::
260 GCGCGCGCATCACCGAGGGGCA.....C 235
   ||| ::::: ||||| ||| ||| |||||:::
222 oProValArgAspProProGlyArgSerGlyProLeuProProProPr 239
   ::::: ||||| ||| ||| ||| |||||:::
234 CGGCTTTTCAGTCAGTCAGCATCAGCGGGCTGGGTCAACGACCCAC 185
   ::::: ||||| ||| ||| ||| |||||:::
239 roValSerArgAsnGlySerThrSerArgAlaLeuProAlaThr..ProG 255
   ::::: ||||| ||| ||| ||| |||||:::
184 CTGCGCCA.....GGTAGCGACTCCGCGCGCAGCAGCGCGCGCC 144
   ||| ::::: ||||| ||| ||| |||||:::
255 InLeuProSerArgSerGlyValAspSerProArgSerGlyProArgPro 271
   ::::: ||||| ||| ||| ||| |||||:::
143 GCGCTGGGCTGATCCACGACGCGGATGTTTCGACAGCGGACTGGT 94
   ||| ||||| |||
272 ProLeuProAspArgPro..... 278
   ::::: ||||| ||| ||| ||| |||||:::
93 GCGGACGAGCCCATCTGCGCGGCTTCCTCGTGGCTGGGTGCGCGCGC 44
   ||||| ||| ||| ||| |||||:::
279 .....SerAlaGlyAlaProProp 285
   ::::: ||||| ||| ||| ||| |||||:::
43 CGGTGCGCGCCACC 30
   || |||||:::
285 roProProProSer 289
   ::::: ||||| ||| ||| ||| |||||:::

```

seq_name: sp_bacteria:Q9XDH2

```

seq_documentation_block:
ID Q9XDH2 PRELIMINARY; PRT: 763 AA.
AC Q9XDH2;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=20090472; PubMed=10627046;
RA Espitia C.; Lacleite J.P.; Mondragon-Palomino M.; Amador A.;
RA Campuzano J.; Martens A.; Singh M.; Cicero R.; Zhang Y.; Moreno C.;
RT "The PE-PGR glycine-rich proteins of Mycobacterium tuberculosis: a
  new family of fibronectin-binding proteins?";
RL Microbiology 145:3487-3495(1999).
DR EMBL; AF071081; RAD41594.1; -.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistil_extensin.
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01217; PRICHEXTNSN.
DR PRINTS; PR01218; PSTLEXTENSIN.
SQ SEQUENCE 763 AA; 75035 MW; 39168EC45A5916F8.CRC64;

```

alignment_scores:
 Quality: 134.00 Length: 138
 Ratio: 1.887 Gaps: 8
 Percent Similarity: 51.449 Percent Identity: 36.957

Seq ID	Seq Name	Accession	Length	Quality	Ratio	Percent Similarity	Percent Identity	Gaps
OC	Chlamydomonadaceae; Chlamydomonas.							
OX	NCBI_TaxID=3055;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RA	Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J., Goodenough U.W.;							
RT	"Glycosylated polyproline II rods-with-kinks as a structural motif in plant hydroxyproline-rich glycoproteins."							
RL	Biochemistry 0:0-0(2001).							
DR	EMBL; AF309494; AAG45420.1;							
DR	InterPro; IPR003882; Pistil_extensin.							
DR	InterPro; IPR002965; P_rich_extensn.							
DR	PRINTS; PR01217; PRICHEXTENS.							
DR	PRINTS; PR01218; PSTLTEXTENSIN.							
DR	SEQUENCE 555 AA; 54219 MW; 6A58A90465502F5 CRC64;							
SO	SEQUENCE							
alignment_scores:								
Quality:	130.50		148					
Ratio:	2.008		9					
Percent Similarity:	43.919		35.135					
alignment_block:								
US-09-462-480-3/rev x Q9FPQ6								
Align seg 1/1 to: Q9FPQ6 from: 1 to: 555								
469	CCCCGGTGGCGGGAAGTCTGTTGTCATTACGGGAGCTCACCAGTCGGTCC	420						
68	ProGlyProProSer.....Pr 73							
419	TCTTGTGTCAGTCGTCTGCTGCTTTCACGCTCTCTCGCGAGCGG	370						
73	oAlaProProSerProPro.....SerProAlaProProS 85							
369	TGCCGGCGGACACAGACCGGGCTGGTGGACCGCGGATGCCAACCT	320						
85	erProAlaProProSerProAlaProProSerProAlaProProSerPro 101							
319	GGCCCATCGCTCCGACCCACCGAGGGCGCCACCGCTCACCAGCAT	270						
102	AlaProProSerProAlaProProSerProAlaProProSerProAla.. 117							
269	CCGGCAACAGCCCGGCGATCACCAGGGGGCAACCGGCTTTTCGATCAG	220						
118ProProSerProProSerProAlaProPro..SerProSe 130							
219	CTGAGACATCAGCGGCTGGGTGTCACGACCCACCTCGCGCAGGTAGCG	170						
130	r.....ProProAlaProProSerP 137							
169	ACTCGCGCGCAGCAGCGCGCGCGGCTGGGGCGCTGATCCA...CCA 123							
137	roSerProProSer....ProAlaProProLeuProProSerProAlaPro 152							
122	GCCAGCGGATGTTTCGACAGCGGACTGGTGGCGGACGAGCGCCATCTGGCG	73						
153	ProSer.....ProSerProProValProPr 161							
72	GGTTCCTCTGCGCTGGGTGGCGCGCGCGGTCGCCGCCACC	30						
161	oSerProSer.....ProProValProProSer 170							
seq_name:	sp_mammal:Q9N1P0							
seq_documentation_block:								
ID	Q9N1P0							
AC	Q9N1P0; PRELIMINARY; PRT; 818 AA.							
DT	01-OCT-2000 (TrEMBLrel. 15, Created)							
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)							
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)							
GN	SUBMAXILLARY MUCIN (FRAGMENT).							
DN	BSM1.							

RP	SEQUENCE FROM N.A.
RX	MEDLINE=98153262; PubMed=9480768;
RA	Hayashi C.Y., Lewis R.V.;
RT	"Evidence from flagelliform silk cDNA for the structural basis of
RT	elasticity and modular nature of spider silks.";
RL	J. Mol. Biol. 275:773-784(1998).
DR	EMBL; AF027972; AAC38846.1; -
DR	InterPro; IPR000087; Collagen.
DR	InterPro; IPR00209; Peptidase S8.

```

seq_documentation_block:
ID   Q9WUE8      PRELIMINARY;      PRT;   2087 AA.
AC   Q9WUE8;
DT   01-NOV-1999 (TREMBLrel. 12, Created)
DT   01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE   SHANKIA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=99360650; PubMed=10433258;
RA   Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtchanoff J.,
RA   Weinberg R.J., Worley P.F., Sheng M.;
RT   "Shank, a novel family of postsynaptic density proteins that binds to
RT   the NMDA receptor/PSD-95/GKAP complex and cortactin.";
RL   Neuron 23:569-582(1999).
DR   EMBL; AF131951; AAD29417.1; -.
DR   HSP; P00519; IABL.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR001478; PDZ.

```

DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR001660; SAM.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00536; SAM; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR SMART: SM00248; ANK; 3.
DR SMART: SM00228; PDZ; 1.
DR SMART: SM00454; SAM; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00088; ANK_REPEAT; 3.
DR PROSITE: PS0297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS0106; PDZ; 1.
DR PROSITE: PS00002; SH3; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 2087 AA; 218125 MW; 0955C33D7A7AA8F8 CRC64;

alignment_scores:
Quality: 127.50 Length: 124
Ratio: 1.903 Gaps: 9
Percent Similarity: 54.032 Percent Identity: 36.290

alignment_block:

US-09-462-480-3/rev x Q9WUE8

Align seg 1/1 to: Q9WUE8 from: 1 to: 2087

353 CCGGG.....CTGGTGAGCCGCCGAT...TCGGAACCTG 319
||||| ||| :|||: |||
1512 ProGlyProHisProLeuProAspProSerProAlaThrProLe 1528

318 GCCATCGCTCCCGACCC...ACGGAGCGCGCCACCTCCACCGAG 272
||| ||||| ||| :|||: |||
1528 uProAlaAlaProProAlaValAlaAlaProProThrLeuAsps 1545

271 ATCCGGCAACAGCC.....GCCGGCATCACC 246
: |||: |||
1545 erThAlaSerSerLeuThrSerTyrAspSerGluValAlaThrLeuThr 1561

245 GAGGGGCAACCGCTTTTCATCAGTCAGACATCAGCGCGTGGGGT 196
: |||||
1562 GlnGlyAlaPro.....AlaAlaPr 1568

195 CAACGACCCCTCGCGAGT.....AGGACTCGCGCGCAGCAGC 152
: ||||| ||| :|||: |||
1568 oGlyAspProAlaProGlyProProAlaProAlaProAlaProp 1585

151 CCGCGCCCGCTGGGCTGATCCACGACGCGGATGTTCCAGACAGC 102
||||| ||| ||||| |||
1585 roAlaProGlnProGlyProAspProProGly.....ThrAspSer 1599

101 GGACTG.....GTCCGAGCAGG.....CCCATCTGCGC 73
|||: ||| ||||| |||: |||
1600 GlyIleGluValAspSerArgSerSerSerAspHisProLeuGluTh 1616

72 GGCTTCCTCGTGGGTG 51
: |||||: |||
1616 rIleSerSerAlaSerThrLeu 1623

seq_name: sp_rodent:Q9WU13

seq_documentation_block:

ID Q9WU13 PRELIMINARY; PRT; 2158 AA.
AC Q9WU13;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SYNAPTIC SAPAP-INTERACTING PROTEIN SYNAMON.
OS Rattus norvegicus (Rat)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99419021; PubMed=10488079;
RA Yao I., Hata Y., Hiraio K., Deguchi M., Ide N., Takeuchi M., Takai Y.;
RT "Synamon, a novel neuronal protein interacting with synapse-associated
protein 90/postsynaptic density-95-associated protein.";
RL J. Biol. Chem. 274:27463-27466(1999).
DR EMBL: AF102855; AAD04569.2;
DR HSSP: P00519; IABL
DR InterPro: IPR002110; ANK.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR001660; SAM.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00536; SAM; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR SMART: SM00248; ANK; 3.
DR SMART: SM00228; PDZ; 1.
DR SMART: SM00454; SAM; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00088; ANK_REPEAT; 3.
DR PROSITE: PS0297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS0106; PDZ; 1.
DR PROSITE: PS00002; SH3; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 2158 AA; 225519 MW; 2AB6D53B5F1A4407 CRC64;

alignment_scores:

Quality: 127.50 Length: 124
Ratio: 1.903 Gaps: 9
Percent Similarity: 54.032 Percent Identity: 36.290

alignment_block:

US-09-462-480-3/rev x Q9WU13

Align seg 1/1 to: Q9WU13 from: 1 to: 2158

353 CCGGG.....CTGGTGAGCCGCCGAT...TCGGAACCTG 319
||||| ||| :|||: |||
1583 ProGlyProHisProLeuProAspProSerProAlaThrProLe 1599

318 GCCATCGCTCCCGACCC...ACGGAGCGCGCCACCTCCACCGAG 272
||| ||||| ||| :|||: |||
1599 uProAlaAlaProProAlaValAlaAlaProProThrLeuAsps 1616

271 ATCCGGCAACAGCC.....GCCGGCATCACC 246
: |||: |||
1616 erThAlaSerSerLeuThrSerTyrAspSerGluValAlaThrLeuThr 1632

245 GAGGGGCAACCGCTTTTCATCAGTCAGACATCAGCGCGTGGGGT 196
: |||||
1633 GlnGlyAlaPro.....AlaAlaPr 1639

195 CAACGACCCCTCGCGAGT.....AGGACTCGCGCGCAGCAGC 152
: ||||| ||| :|||: |||
1639 oGlyAspProAlaProGlyProProAlaProAlaProAlaProp 1656

151 CCGCGCCCGCTGGGCTGATCCACGACGCGGATGTTCCAGACAGC 102
||||| ||| ||||| |||
1656 roAlaProGlnProGlyProAspProProGly.....ThrAspSer 1670

101 GGACTG.....GTCCGAGCAGG.....CCCATCTGCGC 73
|||: ||| ||||| |||: |||
1671 GlyIleGluValAspSerArgSerSerSerAspHisProLeuGluTh 1687

72 GGCTTCCTCGTGGGTG 51
: |||||: |||

[illegible]

OM of: US-09-462-480-2 to: A_Geneseq_032802:* out_format : pfs
Date: Jul 22, 2002 1:22 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ntp.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US09462480/runat_18072002_164418_19413/app_query.fasta_1.2850
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40 cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZB=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480 -CGN1_1_57 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-2
Query length: 524
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 134.850000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32452 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32384 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81705 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64338 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39135 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW38992 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW19844 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72949 +	713.00	1059.72	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW21946 +	713.00	1059.72	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72948 +	373.00	559.42	3.7e-23		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW21945 +	373.00	559.42	3.7e-23		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAE01594 +	147.50	214.74	0.0003		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AAW54248 +	145.50	215.70	0.0004		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW79841 +	141.50	194.83	0.0010		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW24091 +	140.00	199.36	0.0012		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG21932 +	136.50	193.72	0.0023		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW31855 +	136.50	190.49	0.0024		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG13470 +	135.50	200.47	0.0025		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW79128 +	135.50	198.27	0.0026		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG14738 +	133.00	185.05	0.0047		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW20852 +	132.50	196.20	0.0044		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72204 +	132.50	172.16	0.0062		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW39141 +	131.00	182.93	0.0069		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG14734 +	128.50	175.80	0.0117		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG21919 +	127.50	184.50	0.0122		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG30150 +	127.50	182.36	0.0126		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG31516 +	127.50	168.19	0.0154		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AAW12000 +	127.50	168.18	0.0154		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW18664 +	126.50	181.50	0.0151		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:AAW60620 +	125.50	165.76	0.0224		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG03626 +	125.00	183.38	0.0190		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABW70063 +	124.50	177.24	0.0226		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABW11016 +	124.00	174.78	0.0255		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW29514 +	123.50	173.21	0.0284		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AAW44463 +	123.50	173.21	0.0284		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW13804 +	123.50	173.21	0.0284		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW18528 +	123.00	179.52	0.0283		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG14000 +	123.00	172.42	0.0313		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABW65293 +	122.50	165.62	0.0375		

/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABW61602 - 122.00 171.75 0.0
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG04558 - 121.00 176.14 0.0
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG19764 - 121.00 174.96 0.0
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW68408 - 121.00 172.86 0.0
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:AAW28150 - 121.00 172.14 0.0

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32452

seq_documentation_block:

ID AAW32452 standard; Protein; 368 AA.

AC AAW32452;

DT 09-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb37-FL.

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M.tuberculosis.

OS Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT Misc-difference 10

FT /note= "Any amino acid"

PN WO9709428-A2.

PD 13-MAR-1997.

PF 30-AUG-1996; 96WO-US14674.

PR 12-JUL-1996; 96US-0680574.

PR 01-SEP-1995; 95US-0523436.

PR 22-SEP-1995; 95US-0533634.

PR 02-MAR-1996; 96US-0620874.

PR 05-JUN-1996; 96US-0659683.

PA (CORI-) CORIXA CORP.

PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI Twardzik DR, Vedvick TH;

XX WPI; 1997-192903/17.

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are useful in vaccines for prevention or treatment of tuberculosis, also for diagnosis

PS Example 3; Page 146-147; 168pp; English.

CC A new immunogenic polypeptide has been developed comprising an immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its variant differing only in conservative substitutions and/or modifications). The present sequence represents a M.tuberculosis antigen, Tb37-FL. The immunogenic protein, and fusion proteins containing one or more of the proteins or one of the proteins plus ESAT-6, are useful in vaccines, preferably when formulated with a non-specific adjuvant, to induce an immune response against M.tuberculosis (for treatment or prevention).

Sequence 368 AA;

alignment_scores: Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-2 x AAW32452 ..

Align seg 1/1 to: AAW32452 from: 1 to: 368

1 CTGCAGCAGGTGACCTGCTGTTGTTACGAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG1 242
51 CAACCCAGCGGACGAGGAGCGCGCAGATGGCGCTGCTCGGCACCATC 100
242 yasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCGCTGGTGGTGCATCAGGCCCGCGCGCGCG 150
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTCGCGCGGAGTCTGCTGCGGAGTGGGTGGTGGTGGTGGT 200
276 GlyLeuLeuArgAlaGlnSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTCATCGAAAGCCGCTGCGCCCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuLeuGlyLysProValAlaProSerValM 309
251 TGCCGCGCGCTGTTCCGCGATCGTGGTGGTGGTGGTGGTGGTGGT 300
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGCGGAGCGTGGCGAGGTTCGCAATCCGCGCGCTCCACGAGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTGTGTCGCGCGCGGCGCGCTGCGCGAGGCGTGAAGAAGACGACG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGAGCGTGGGACGAGGAGGAGGACTGG 429
359 luAspAspTrpAspGluAspAspTrp 368

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32384

seq_documentation_block:

ID AAW32384 standard; Protein: 368 AA.

AC AAW32384;

DT 13-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb37-FL.

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M.tuberculosis.

OS Mycobacterium tuberculosis.

PH Key Location/Qualifiers

FT Misc-difference 10

FT /note= "Any amino acid"

PN wO9709429-A2.

PD 13-MAR-1997.

PF 30-AUG-1996; 96WO-0514675.

PR 12-JUL-1996; 96US-0680573.

PR 01-SEP-1995; 95US-0523435.

PR 22-SEP-1995; 95US-0532136.

PR 22-MAR-1996; 96US-0620280.

PR 05-JUN-1996; 96US-0658800.

XX (CORI-) CORIXA CORP.

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI Twardzik DR, Vedvick TH;

PI

XX WPI; 1997-192904/17.
XX DR
XX PT
XX - useful for diagnosis of M. tuberculosis infection
XX PS
XX Example 3; Page 159-161; 190pp; English.
XX CC
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC its variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis
CC antigen, Tb37-FL. The immunogenic polypeptide can be used to diagnose
CC M.tuberculosis infection by forming complexes with specific
CC antibodies in the sample. Fragments of DNA encoding the immunogenic
CC polypeptide can be used as diagnostic primers or probes and agents
CC that bind to the antigen, especially monoclonal antibodies or
CC equivalent polyclonal antibodies, are also used for diagnosis.
XX SQ Sequence 368 AA;

alignment_scores:

Quality: 713.00 Length: 143

Ratio: 5.057 Gaps: 0

Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-2 x AAW32384

Align seg 1/1 to: AAW32384 from: 1 to: 368

1 CTGCAGCAGGTGACCTGCTGTTGTTACGAGGTGGCGGCACCGCGCGG 50

|||||

226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG1 242

51 CAACCCAGCGGACGAGGAGCGCGCAGATGGCGCTGCTCGGCACCATC 100

242 yasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTCGAACCATCGCTGGTGGTGCATCAGGCCCGCGCGCGCG 150

|||||

259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCCTGTCGCGCGGAGTCTGCTGCGGAGTGGTGGTGGTGGTGGTGGT 200

|||||

276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292

201 CACGCCGCTGATGCTCAGCTCATCGAAAGCCGCTGCGCCCTCGGTGA 250

292 gThrProLeuMetSerGlnLeuLeuGlyLysProValAlaProSerValM 309

251 TGCCGCGCGCTGTTCCGCGATCGTGGTGGTGGTGGTGGTGGTGGT 300

|||||

309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325

301 GGTCCGCGGAGCATGGCGCAGGTTTCGCAATCCGCGCGCTCCACGAGCC 350

|||||

326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342

351 GGGTGTGTCGCGCGGCGCGCTGCGCGAGGCGTGAAGAAGACGACG 400

|||||

342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359

401 AGGAGCGTGGGACGAGGAGGAGGACTGG 429

359 luAspAspTrpAspGluAspAspTrp 368

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81705

seq_documentation_block:

ID AAW81705 standard; Protein: 368 AA.

XX


```

AC AAW81705;
XX
XX 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide Tb37-FL.
XX
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX KW vaccine; pharmaceutical; infection; diagnosis.
XX
XX Mycobacterium tuberculosis.
XX
XX Key Location/Qualifiers
FH Misc-difference 10
FT /label= unknown
FT
XX
XX WO9816646-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18293.
XX
XX 13-MAR-1997; 97US-0818112.
XX PR 11-OCT-1996; 96US-0730510.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX WPI; 1998-261042/23.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
XX to develop products for the detection of M. tuberculosis infection
XX and for diagnosis, treatment and prevention of tuberculosis
XX
XX Example 3B; Page 137-138; 230pp; English.
XX
XX This sequence represents an immunogenic portion of a soluble
XX Mycobacterium tuberculosis (MT) antigen which can be used in a method
XX for inducing protective immunity against tuberculosis (TB). This
XX sequence can be formulated into vaccines and/or pharmaceutical
XX compositions for immunising against M. tuberculosis infection or may
XX be used for the diagnosis of tuberculosis.
XX
XX Sequence 368 AA;

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x AAW81705 ..

Align seg 1/1 to: AAW81705 from: 1 to: 368

1 CTGCAGCAGGTGACGTGCTTCTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
|||||
51 CAACCCAGCCGACGAGGAAGCCGCGCAGATGGGCTGCTCGGCACCAAGTC 100
|||||
242 YAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACATCCGCTGCTGGTGGATCAGCCCGCCAGCGCGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGTGTCGCGCGGAGTCCCTACTCGCGCAGGTGGTGGTGCACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292

```

```

201 CACGCCGCTGATGTCTCAGCTGATCGAAAGCCGGTTGCCCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TCCCGCGCGCTGTTCCGGATCGTTCGGTGACGGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCGCGGAGCGATGGCGGTTTCGCAATCCGCGCTCCACACAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTCGCGCGCCGACCGCTCGCGCAGGACGCTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGACGAAAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:AAW64338
seq_documentation_block:
ID AAW64338 standard; Protein; 368 AA.
XX
XX AAW64338;
XX
XX 09-NOV-1998 (first entry)
XX
XX Mycobacterium tuberculosis antigen Tb37-FL.
XX
XX Tuberculosis; infection; diagnosis; antigen; Tb37-FL.
XX
XX Mycobacterium tuberculosis strain H37Rv.
XX
XX Key Location/Qualifiers
FH Misc-difference 10 /note= "unidentified"
FT
FT WO9816645-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18214.
XX
XX 13-MAR-1997; 97US-0818111.
XX PR 11-OCT-1996; 96US-0729622.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX WPI; 1998-251292/22.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX to develop products for the detection of M. tuberculosis infection
XX and diagnosis of tuberculosis
XX
XX Example 3; Page 143-144; 250pp; English.
XX
XX This polypeptide comprises Mycobacterium tuberculosis antigen
XX Tb37-FL. It is encoded by genomic DNA isolated from a M.
XX tuberculosis strain H37Rv genomic library using a probe from
XX clone Tb38-1 (see AAV44384). The invention relates to compositions
XX and methods for diagnosing tuberculosis. It provides polypeptides
XX (see AAW64291-W64379) comprising an antigenic portion of a soluble
XX M. tuberculosis antigen, or an immunogenic portion of an M.
XX tuberculosis antigen, as well as DNA sequences encoding such
XX polypeptides, recombinant expression vectors and transformed or
XX kits for detecting M. tuberculosis infection in a patient using
XX these polypeptides, antibodies or oligonucleotide probes and

```

CC primers, for the diagnosis of tuberculosis.

XX
SQ Sequence 368 AA;

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-2 x AAW64338 ..

Align seg 1/1 to: AAW64338 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGCTGTTTCAGCCAGGTGGCGGCACCGCGGGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCTGCTCGGCACCAATC 100
|||||
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCAACCATCCGCTGCTGCTGATCAGCCCGCCAGCGGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCTGTCTGCGCGGAGTCTGCTACCTGCGCGCAGGTGGTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
201 CACGCCGCTGATGTCATCAGTCGATCGAAAGCCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTTCGCGGATCGCTGCTGAGCGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCGGGAGCGAGTGGCGGCTGCAATCCGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGTCGCGCGGACCGCTCGCGCAGGAGCGGTGAAGAACGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGACGCTGGGAGAGGAGGACGACTGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV39135.

seq_documentation_block:

ID AAV39135 standard; Protein: 368 AA.

XX AC AAV39135;

XX 05-NOV-1999 (first entry)

DE M. tuberculosis antigen Tb37-FL amino acid sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.

XX Mycobacterium tuberculosis.

OS WO9942076-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03268.

XX 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX (CORI-) CORIXA CORP.
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1999-527409/44.
XX New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX Example 3; Page 132-133; 299pp; English.
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA19249 to AA219460 and AAV39083 to
CC AAV39225 are used in the exemplification of the present invention.
XX Sequence 368 AA;
SQ

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-2 x AAV39135 ..

Align seg 1/1 to: AAV39135 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGCTGTTTCAGCCAGGTGGCGGCACCGCGGGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCTGCTCGGCACCAATC 100
|||||
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCAACCATCCGCTGCTGCTGATCAGCCCGCCAGCGGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCTGTCTGCGCGGAGTGGCGGCTGCAATCCGCGGCTCCACCGCC 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
201 CACGCCGCTGATGTCATCAGTCGATCGAAAGCCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTTCGCGGATCGCTGCTGAGCGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCGGGAGCGAGTGGCGGCTGCAATCCGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGTCGCGCGGACCGCTCGCGCAGGAGCGGTGAAGAACGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359

401 AGGACGACTGGGACGAAGAGGACGACTGG 429
|||||
359 luAspAspTTPAspGluAspAspTrp 368

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AA1999

seq_documentation_block:
ID AAY38992 standard; Protein; 368 AA.

AC AAY38992;
XX
DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen protein Tb37-FL.
XX
KW Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.

XX Mycobacterium tuberculosis.
XX
PN W09942118-A2.
XX
PD 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03265.
XX
PR 05-MAY-1998; 98US-0072596.
PR 18-FEB-1998; 98US-0024753.

XX (CORI-) CORIXA CORP.
PA
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI: 1999-527416/44.

XX New polypeptide comprising antigenic portions of M. tuberculosis
PT
XX Example 3; Page 177-179; 323pp; English.
XX
XX This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.
XX
XX Sequence 368 AA;

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x AAY38992 ..
Align seg 1/1 to: AAY38992 from: 1 to: 368

1 CTGACGAGGTGACGTCGTTCTTCAGCCAGGTGGCGGACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
|||||
51 CAACCCAGCCGACGAGGAGCGGCGGATGGCGCTGCTCGGACCGAGTC 100
|||||
242 YasnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCGAACATCCGCTGCTGGTGTGATCAGGCCCGCGGCGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||

151 GGCCTGCTGGCGCGGAGTGCTACCTGCGGAGGTGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CAGCCGCTGATGTCCTCAGTGATCGAAAGCCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerVal 309
|||||
251 TGC CGCGGCTGTGTCGCGGATCGGTCGTCGAGCGGTGGCGGCTCC 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGGAGCGATGGCCAGGTTCCGAATCCGGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
|||||
351 GGTCTGCTGGTGGCGCGGCGGCTCGGCGGAGGCGGTGAAGAACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
|||||
401 AGGACGACTGGGACGAAGAGGACGACTGG 429
|||||
359 luAspAspTTPAspGluAspAspTrp 368

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AA19844

seq_documentation_block:
ID AAB19844 standard; Protein; 368 AA.

XX AAB19844;
XX
DT 05-MAR-2001 (first entry)
XX
DE Mycobacterium tuberculosis protein MTBN3.

XX MTBN3; tuberculosis; BCG; vaccine; infection; diagnosis.
XX
XX Mycobacterium tuberculosis.

XX WO2000066157-A1.

XX 09-NOV-2000.

XX 04-MAY-2000; 2000WO-US12257.

XX 04-MAY-1999; 99US-0132505.

XX (PUBL-) PUBLIC HEALTH RES INST NEW YORK.

XX Gennaro ML;

XX WPI: 2001-007153/01.

XX N-PSDB; AAA89037.

XX Novel polypeptide encoded by open reading frames present in
PT Mycobacterium tuberculosis genome and not by the BCG strain of M.
PT bovis, useful as vaccine and for diagnosing tuberculosis infection
XX
PS Claim 11; Fig 1; 35pp; English.

XX The present sequence is that of the Mycobacterium tuberculosis
CC MTBN3 protein. This is 1 of 8 proteins, i.e. MTBN1-8 (see
CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)
CC identified as being present in the genome of M. tuberculosis but
CC absent from the genome of the BCG strain of Mycobacterium bovis.
CC MTBN1-8 represent reagents that are useful in discriminating between
CC M. tuberculosis and BCG and, in particular, for diagnostic methods
CC which discriminate between exposure of a subject to M. tuberculosis
CC and vaccination with BCG. The invention features these MTBN
CC polypeptides, functional fragments of them, DNA encoding them,
CC vectors, transformed cells, and diagnostic, therapeutic, and
CC prophylactic (vaccine) methods, including genetic vaccination
CC methods.

362 luAspAspTrpAspGluAspAspTrp 371

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV21946

seq_documentation_block:

ID_AAY21946 standard; Protein; 371 AA.

AC AAY21946;

DT 06-SEP-1999 (first entry)

DE Amino acid sequence of antigen RD1-ORF5.

XX Immunogenic; Mycobacterium tuberculosis; immune response; infection;
KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;
KW pharmacological; vaccination; M. africanum; M. bovis; CFP7A; CFP30A;
KW CFP7B; CFP19; CFP27; CFP30A; RD1-ORF; CFP10A; CFP16; CFP19; CFP23;
KW CFP25A; CFP30B; CFP7B.

XX Mycobacterium tuberculosis.

OS WO9924577-Al.

PN 20-MAY-1999.

PD 08-OCT-1998; 98WO-DK00438.

PF 01-APR-1998; 98WO-DK00132.

PR 10-NOV-1997; 97DK-000277.

PR 05-JAN-1998; 98US-0007488.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Skjot R;

XX WPI: 1999-347282/29.

DR N-PSDB; AAX81046.

XX New immunogenic fragment of Mycobacterium tuberculosis

XX Example 2; Page 219-220; 265pp; English.

CC The invention describes a substantially pure immunogenic polypeptide
CC fragment (I) from Mycobacterium tuberculosis that is able to evoke a
CC protective immune response against infections by mycobacteria belonging
CC to the tuberculosis complex. The invention provides a (1) fusion
CC polypeptide comprising at least one polypeptide fragment (I) and at least
CC one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell
CC epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second
CC different amino acid sequence from M. tuberculosis, and/or including a
CC sequence which protects the first amino acid sequence from in vivo
CC degradation or post-translational processing; (3) a nucleic acid fragment
CC that encodes the above polypeptides. The polypeptides and nucleic acid
CC are useful as pharmaceuticals, for diagnosis of and as antigens for
CC vaccination against TB caused by Mycobacterium tuberculosis, africanum or
CC bovis. The polypeptides are also useful for diagnosing ongoing or
CC previous sensitization in an animal with bacteria belonging to the
CC tuberculosis complex. The invention also describes the use of CFP7A or
CC CFP30A or a T-cell epitope of for the induction of a strong immune
CC response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell
CC epitope of for diagnosis of TB in a mammal by performing a DTH type skin
CC test; use of CFP27, CFP30A, RD1-ORF2, RD1-ORF3, RD1-ORF5, MPT59-ESAT6,
CC ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-
CC cell epitope of for the preparation of an immunological composition; and
CC for the preparation of a subunit vaccine.

XX Sequence 371 AA;

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-2 x AAY21946

Align seg 1/1 to: AAY21946 from: 1 to: 371

1 CTGCAGAGGTGACGTCGTTGTTTCAGCCAGGTGGCGGCACCCGGCGGG 50
|||||
229 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 245
51 CAACCCAGCCGACGAGGAGGAGCCGCGGAGATGGGCGCTGCTGGCAGCAGTC 100
|||||
245 YAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 262
101 CGCTGTCTGAACCATCCGCTGGTGGATCAGCCGCCAGCGCGGGCGG 150
|||||
262 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 278
151 GGCCTGCTGCGCGGAGTGCCTACCTGGCGCAGGTGGGTGCTTGCACCCG 200
|||||
279 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 295
201 CACGCCCTGATGTCCTCAGCTGATCGAAACCGGTTGCCCGCTCGGTGA 250
|||||
295 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValm 312
251 TGGCGGCGGCTGTTGCGGATCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 300
|||||
312 etProAlaAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 328
301 GGTCCGGGAGCGATGGCCAGGGTTTCGCAATCCGCGGCTCCACCGAGCC 350
||| |||||
329 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 345
351 GGGTCTGGTCGCGCGGCGCACCGCTCGCGCAGGAGCGGTGAAGAGACGACG 400
|||||
345 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 362
401 AGGACGACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 429
|||||
362 luAspAspTrpAspGluAspAspTrp 371

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72928

seq_documentation_block:

ID_AAW72928 standard; Protein; 139 AA.

AC AAW72928;

DT 21-JAN-1999 (first entry)

DE Mycobacterium tuberculosis antigen RD1-ORF4.

XX Mycobacterium tuberculosis; antigen; vaccine; immunological;

XX immunogen; infection.

XX Mycobacterium tuberculosis.

PN WO9844119-Al.

PD 08-OCT-1998.

PF 01-APR-1998; 98WO-DK00132.

PR 05-JAN-1998; 98US-0070488.

PR 02-APR-1997; 97DK-0000376.

PR 18-APR-1997; 97US-0044624.

PR 10-NOV-1997; 97DK-0001277.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB;

PI Rosenkrands I, Weidigh K;

67 s 67

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAE01594
seq_documentation_block:
ID  AAE01594 standard; Protein; 314 AA.
XX
AC  AAE01594;
XX
DT  17-JUL-2001 (first entry)
XX
DE  Human gene 12 encoded secreted protein HHSW69, SEQ ID NO:144.
XX
KW  Human; secreted protein; proliferative disorder; cancer; tumour;
KW  foetal abnormality; developmental abnormality; haematopoietic disorder;
KW  immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW  inflammation; allergy; neurological disorder; Alzheimer's disease;
KW  Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW  skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW  cardiovascular disorder; angiogenic disorder; kidney disorder;
KW  gastrointestinal disorder; pregnancy-related disorder;
KW  endocrine disorder; infection; wound healing; vulnerability;
KW  cell culture; chemotaxis; food additive; chromosome 17;
KW  binding partner identification; gene therapy.
XX
OS  Homo sapiens.
XX
FH  Key Location/Qualifiers
FT  Peptide 1..26
FT  /label= Signal_peptide
FT  Protein 27..55
FT  /label= Mature_human_secreted_protein
FT  Misc-difference 129
FT  /label= Unknown
FT  /note= "Encoded by MCG"
XX
PN WO200134623-A1.
XX
PD 17-MAY-2001.
XX
PF 01-NOV-2000; 2000WO-US30037.
XX
PR 05-NOV-1999; 99US-0163577.
PR 30-JUN-2000; 2000US-0215137.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Moore PA;
XX
DR WPI; 2001-316490/33.
DR N-PSDB; AAD05437.
XX
PT Nucleic acids encoding 29 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Claim 11; Page 496-497; 535pp; English.
XX
CC AAD05389-AAD05473 represent cDNAs corresponding to 29 human secreted
CC protein genes, and AAE01546-AAE01630 represent the proteins they encode.
CC AAE01631-AAE01660 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 29 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
```

```
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
XX secreted protein of the invention.
SQ Sequence 314 AA;
```

```
alignment_scores:
Quality: 147.50 Length: 163
Ratio: 2.049 Gaps: 10
Percent Similarity: 44.172 Percent Identity: 33.742

alignment_block:
US-09-462-480-2/rev x AAE01594 ..
Align seg 1/1 to: AAE01594 from: 1 to: 314
472 CGGCGCGGGTGGCGGGAAGTCTGTTCATACGGGAGCTCACCAGTCG 423
|||||
126 ArgProGly***Pro..... 130
422 TCCTCTTCGTCAGTCGT.....CCTC 400
|||||
131 .ProLeuProProSerSerSerGlyAsnAspGluThrProArgLeuProG 147
399 GTCGTCTTCTTCACGCTCTCGCGAGCGGTGCGCGCGGACGACACCG 350
|||||
147 InArgAsnLeu...SerLeuSerSerThrProProLeuProSerPro 162
349 GGCTGGTGGAGCCCGGATTGCGAACCTGGCCCATCGCTCCCGGACCC 300
|||||
163 Gly.....ArgSerGlyPro...LeuProProPr 171
299 ACGGAGCGGCGCACCCGTCACGACGATCCGGCAACAGCGCGCGCAT 250
|||||
171 OProSerGluArg...ProProProValArgAspProProGlyArgS 187
249 CA.....CCGAGGGGGCAACCGCTTTTCGATCAGTCAGTACATCA 209
|||||
187 erGlyProLeuProProProValSerArgAsnGlySerThrSer 203
208 GCGCGTGGGGTCAACGACCCACCTCGGCA.....GGTAGCGAC 168
|||||
204 ArgAlaLeuProAlaThr...ProGlnLeuProSerArgSerGlyValAsp 219
167 TCGCGCGGACGAGCGCGCGCGCTGGGGCTGTATCCACCAGCCAG 118
|||||
220 SerProArgSerGlyProArgProProLeuProProAspArgPro..... 234
117 CGGATGTTTCACAGCGGAGCTGTGCGGAGCAGGCCCATCTGCGCGGCTT 68
234 ..... 234
67 CCTCGTGGGTGGTTCGCCCGCGCGGTGGCGCCACC 30
235 .....SerAlaGlyAlaProProProProSer 245
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AAE54248
seq_documentation_block:
ID AAB54248 standard; Protein; 200 AA.
XX
```



```
alignment_block:
US-09-462-480-2/rev x AAY24091
Align seg 1/1 to: AAY24091 from: 1 to: 504

472 CGGCGCGGGTGGCGGGGAAGTCTGTTGTCATTACGGGAGCTCACCAGTCG 423
|||||
315 ArgProGlyProPro..... 319
422 TCCTCTTCGTCCTCCAGTCGT.....CCTC 400
|||||
320 .ProLeuProSerSerSerGlyAsnAspGluThrProArgLeuProG 336
399 GTCGTCCTTCCTTCACGCTCCTCGCGAGCGGTGCGGGCGCGACAGACCG 350
:|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
336 lnArgAsnLeu...SerLeuSerSerThrProLeuProSerPro 351
349 GGCTGGTGGAGCCCGCGGATTGCCAACCTGGCCCATCGTCCCGGACCC 300
||| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
352 Gly.....ArgSerGlyPro...LeuProProPr 360
299 ACCGAGCGGGCCACCGTCACCGACGATCCGGCACAGCGCGCGCAT 250
| :|:|:|:|:| :|:| :|:|:|:| :|:| :|:|
360 ovalProSerGluArgProProProValArgAspProGlyArgS 377
249 CA.....CCGAGGGGCAACCGGCTTTCGATCAGCTGACACATCA 209
||| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
377 erGlyProLeuProProProProValSerArgAsnGlySerThrSer 393
208 CGGCGTGGCGGTCAACGCCACCTCGGCCA.....GTAGCGAC 168
||| :|:|:|:| :|:| :|:| :|:| :|:|
394 ArgAlaLeuProAlaThr..ProGlnLeuProSerArgSerGlyValAsp 409
167 TCGCGCGCAGCAGCGCGCGCGCTGGGCGCTGATCCACAGCCAG 118
||| :|:|:|:| :|:| :|:| :|:| :|:|
410 SerProArgSerGlyProArgProProLeuProProAspArgPro..... 424
117 CGGATGTTGACAGCGGACTGTGCGCGAGCAGGCCCATCTGCGCGGCTT 68
424 ..... 424
67 CCTCGTGGTGGTTCGCGCGCGGTGCGGCCACC 30
||||| :|:|:|:| :|:|:|:|
425 .....SerAlaGlyAlaProProProProProSer 435
```

OM of: US-09-462-480-2 to: Issued_Patents_AA:* out_format : pfs
Date: Jul 22, 2002 1:24 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+nt2p.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09462480/runat_18072002.164418.19425/app_query.fasta_1.2850
-DB-Issued_Patents_AA -QPMW-fastan -SUFFIX-rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -GAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=10.000 -GAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS-human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480 -CGN1_118 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-2
Query length: 524
Database: Issued_Patents_AA:*
Database sequences: 231628
Database length: 24425594
Search time (sec): 53.850000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-112-114			713.00	1096.12	3.9e-54
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-111-109			713.00	1096.12	3.9e-54
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-056-556-114			713.00	1096.12	3.9e-54
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-09-010-928B-2			128.00	181.30	0.0017
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-08-710-928B-2			128.00	179.83	0.0018
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-07-945-283-2			125.50	190.90	0.0021
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-09-413-814-78			125.00	175.04	0.0033
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-09-232-468A-8			121.00	176.27	0.0061
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-09-080-897-2			120.50	164.75	0.0087
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-323-735-2			120.50	164.75	0.0087
/cgn2_6/ptodata/2/1aa/6A_COMB.ppt:US-08-899-595-3			120.00	163.47	0.0097
/cgn2_6/ptodata/2/1aa/6A_COMB.ppt:US-08-483-533-41			118.50	173.62	0.0098
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-283-471A-41			118.50	173.62	0.0098
/cgn2_6/ptodata/2/1aa/PCTUS_COMB.ppt:US-09-010-928B-4			118.00	163.91	0.0134
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-483-533-40			116.00	172.56	0.0151
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-283-471A-40			116.00	172.56	0.0151
/cgn2_6/ptodata/2/1aa/6A_COMB.ppt:US-08-586-165-4			115.00	168.08	0.0198
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-07-928-611-13			112.00	166.24	0.0336
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-08-487-811A-13			112.00	166.24	0.0336
/cgn2_6/ptodata/2/1aa/PCTUS_COMB.ppt:US-09-060-694-13			112.00	166.24	0.0336
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-08-056-051-6			112.00	160.91	0.0381
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-07-928-611-22			112.00	160.91	0.0381
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-08-487-811A-22			112.00	160.91	0.0381
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-060-694-22			112.00	160.91	0.0381
/cgn2_6/ptodata/2/1aa/PCTUS_COMB.ppt:US-093-07370-13			112.00	160.91	0.0381
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-09-041-886-23			111.00	150.49	0.0571
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-764-870-7			110.50	157.47	0.0527
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-980-115-7			110.50	157.47	0.0527
/cgn2_6/ptodata/2/1aa/backfiles1.ppt:5168049-3			110.00	166.73	0.0461
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-483-533-26			110.00	163.46	0.0497
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-09-283-471A-26			110.00	163.46	0.0497
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-08-131-365B-38			110.00	149.25	0.0692
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-08-668-123-38			110.00	149.25	0.0692
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-08-728-323A-2			110.00	149.13	0.0694
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-423-890-8			110.00	146.74	0.0733
/cgn2_6/ptodata/2/1aa/backfiles1.ppt:5171843-9			109.50	159.05	0.0598

/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-199-637A-351 - 109.00 156.67 0.0685
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-556-978B-59 - 109.00 155.74 0.0700
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-199-637A-277 + 109.00 154.56 0.0720
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-07-945-283-2 - 109.00 142.61 0.0950
/cgn2_6/ptodata/2/1aa/backfiles1.ppt:5171843-7 + 108.50 162.14 0.0655
seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-112-114

seq_documentation_block:

Sequence 114: Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 368 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-114

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
alignment_block:
US-09-462-480-2 x US-08-818-112-114
Align seg 1/1 to: US-08-818-112-114 from: 1 to: 368

1 CTGACAGAGTGTGCTGTTTTCAGCCAGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyG1 242
51 CAACCCAGCCACGAGGAGCGCGAGATGGCGCTCTCGGCACGATC 100
|||||
242 YasnProAlaAspGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACCATCGCTGCTGTGTGATCAGGCCCGCGCGCGCG 150

```

;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 368 amino acids
;
; TYPE: amino acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
US-08-818-111-109

```

APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998

```
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-114

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
  US-09-462-480-2 x US-09-056-556-114  ..
  Align seg 1/1 to: US-09-056-556-114  from: 1 to: 368

1  CTGCAGCAGGTGAGTCGCTGTTCTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyGly 242

51  CAACCCAGCCGACGAGGAGCGCGGCGAGATGGGCTGTCTGGCCACCACTC 100
|||||
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGTGTGCAACATCCGCTGCTGCTGGTGTGATCAGGCCCGCGCGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCCTGTGCGCGGAGTGCCTACCTGCGCAGGTGGTCTGCACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292

201 CACGCCGTGTATGTCTACGTGATCGAAAGCCGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309

251 TGCCGGCGGCTGTTCGGGATCGTGGTGACGGGTGGCGCGCTCCCGGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325

301 GGTCGGGACGATGGCGCAGGTTCGCAATCCGCGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342

351 GGGTCTGTCGCGCGGCACCGCTCGCGAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359

401 AGGAGCTGGGACGAAGACGACTGG 429
|||||
359 luAspAspTrpAspGluAspAspTrp 368

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-09-010-928B-28

seq_documentation_block:
; Sequence 28, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
```

```
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 745 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-010-928B-28

alignment_scores:
  Quality: 128.00      Length: 132
  Ratio: 1.707         Gaps: 6
  Percent Similarity: 56.818  Percent Identity: 35.606

alignment_block:
  US-09-462-480-2 x US-09-010-928B-28  ..
  Align seg 1/1 to: US-09-010-928B-28  from: 1 to: 745

37  GGCACCGCGCGCGCAACCCAGCCAGCCAGGAGGAGCGCGAGATGGGCT 86
|||||
251 GlyProGlyGlyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyPr 267

87  GCTCGGACCACTCGCTGTCGAACCATCCGCTGGCTGGTGATCAGGCC 136
|||||
267 oGlyGlyTyrglyProGlyGlySerGlyPro.....GlyGlyAlaGlyP 282

137 CCAGCGCGCGCGCGGCTGCTGCGCGGAGTCTACCTACCTGGCGCAGGT 186
|||||
282 roSer...GlyAlaGlyLeuGlyGlyAlaGly.....ProGlyGlyAla 295

187 GGGTGTGTACCGCGCACCGCCCTGATGTCTCAGTGTATCGAAAGCCGGT 236
|||||
296 Gly.....
237 TGCCCGCTCGGTGATCGCGCGGCTGTTGCGGATCGTCTGCGTACGGGTG 286
|||||
297 .....LeuGlyGlyAlaGlyProGlyGlyAlaGlyThrSerG 309

287 GCGCGCGCTCCG.....GTGGTCCGGAGCGATGGCCAGGTTTCGCAA 330
|||||
309 lAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyGlnGlyAspAla 325

331 TCCGGCGGCTCCACGAGCC....GGTCTGGTCTGCGCGCGCGCCGCTCG 376
|||||
326 GlyProGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyValGly 342

377 GCGAGGACGTGAAGAAGACGACGAGGAGTGGGACGAAGAGA 422
|||||
342 gGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyAlaArgGly 357

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-09-010-928B-2
```

10

```
36 aAlaAlaAlaGlyGlyProLeuArgTyrGlyArgAlaValArgAlaArgG 53
52 AACCC.....AGCCGACGAGGAGCCGCCGAGATGGCGCTGCTCGGCAC 95
   |||  ::||| |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
53 lyProArgAspAlaArgGlyAla.....AlaProGlyArgGly 66
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
96 CAGTCCGCTGTCGAACCATCGCTGGTGGTGGATCAGGCCGCCAGCGCG 145
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
67 AlaArgSerAlaThrProGlyAlaGly.....AlaProAlaArgG1 80
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
146 G.....CGCGGCGCTGCTGGCGCG 165
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
80 yHisArgAlaArgAlaProAlaProArgArgArgGlyProAlaAlaArgG 97
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
166 GAGTCGCTACCTGGCGCAGGTGGTGTGACCCGCGCAGCGCTGATGTC 215
   ||  ||
97 ly..... 97

216 TCAGCTGATCGAAAAGCGGTTGCCCTCGGTGATGCCGCGGTGTG 265
   |||  |||  |||  ::|||  ::|||  ::|||  ::|||  ::|||
98 .....GlyArgGlyGlyGlyProArgAlaGlyAlaLe 108

266 CGGATCTCGGTGACGGTGGCGCC.....GCTCCGGTGGTCCG 306
   |||  |||  |||  |||  |||  |||  |||  |||
108 uArgAlaGlyGly.ArgGlyGlyAlaArgGlyProAlaGluGluGlyAla 124

307 GGAGCGATGGCGAGGTTCCCAATCGCGCGCTCCACCACCGCGG... 353
   |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||
125 GlyAlaAlaGly.GlyValArgLeuProAlaAlaProPro.ProGlyArg 140

354 TCTGTCGCGCGCGCGCTCGCGCAGGAGCGTGAAGAAGCAGCAGG 403
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
141 GlyGlyArgAlaAlaArgProAspProGlyLeuArgArgAlaGlyAl 157

404 AGCACTGGGCGAGCAGCAGCTGGTGAAGTCCCGTAATGAC 446
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
157 aAspAlaGlyArgAspAlaArgArgProGluValArgArgAsp 171
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB pep:US-07-945-283-2

```
seq_documentation block:
; Sequence 2, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EP0 and LIT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4011 ext.513
; TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 1958 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-945-283-2

alignment_scores:
  Quality: 125.50      Length: 238
  Ratio: 1.195        Gaps: 16
  Percent Similarity: 44.118      Percent Identity: 31.092

alignment_block:
US-09-462-480-2 x US-07-945-283-2 ..
Align seg 1/1 to: US-07-945-283-2 from: 1 to: 1958

31 GTGGCGCGCACCGGC...GGCGGCAACCCA..... 57
   |||||  |||  |||||  ::|||
1053 ValGlyGlyGluGlyArgGlyGlyGlyProArgValGlyLeuAlaG1 1069

58 .....GCCGACGAGGAGCGCGCAGATGGGCTGCTGGCACCAGTC 100
   |||  |||  |||  ::|||  ::|||  ::|||  ::|||  ::|||
1069 yArgAspAlaAlaGluAlaValGlyArgGlyValLeuGlyHisGlyP 1086

101 CG...CTGTCGAACCATCCG.....CTGGCTGGTGA..... 129
   ::|||  ::|||  |||||  |||||  |||||  |||||  |||||  |||||
1086 roGluArgAlaProGluProValValLeuGlyGlyGlyGlyGlyGly 1102

130 .....TCAGGCCCC...AGCGCGGG 146
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1103 GlyGlyHisGluArgGlySerGlyValArgSerGlyProGluSerGluG1 1119

147 CGCGGCGCTG..... 156
   |||||  ::|||
1119 yAlaAlaLeuAlaProGlyProProValLeuPheValValAlaValAlaV 1136

157 .....CTCGCGCGGAG.....TCG 171
   ::|||  |||||
1136 alAlaValProAlaGluGlyArgAlaGlyGluProLeuValLeuLeuAla 1152

172 CTACCTGGCGCAGGTGGTTCGTTGACCGCACGCGCGCTGTCTCAGCT 221
   ::|||  |||||  ::|||  |||||  ::|||  ::|||  ::|||  ::|||
1153 ValProGlyAlaAlaGly.....ProGlyArgAlaAlaLe 1164

222 GATCGAAAAGCGGTTGCCCTCGGTGATGCCGCGCGCTGTTCGCGGAT 271
   |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
1164 uLeuLeuAlaProLeuGlyArgTrpValArg...AlaGlyGlyGlyGlyA 1180

272 CGTGGTGACGGTGGCGCGCTCGGTGGTGGTCCGCGGAGCG..... 312
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
1180 laGlyValAlaGlyGlyAlaGlyGluAlaGlyLeuGlyAlaGlyAlaGly 1196

313 .....ATGGCCAGGTTCCAAATCCGCGCGCTCCAC.CAGCCCGGTC 355
   |||  |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
1197 LeuGlyAlaGlyAlaGlyLeuGlyAlaGlyGlyAlaGlyGlyProGlyAl 1213

356 TGGTCGCGCGCGCGCTCGCGCAGGAGCGGTGAGAGAGACA..... 398
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
1213 aGlyGluAlaGly.....GlyGlyAlaArgArgArgArgArgArgA 1227

399 .....CGAGGACGACTGGGA 413
   |||||  ::|||
1227 rgTrpAspAspGluAlaGlyLeuLeuGlyProGluArgGlyGlnAlaGly 1243

414 CGA.....AGAGCAGCACTGGTGAAGTCCCGTAA 442
   |||  |||||  |||  ::|||
1244 ArgGlyLeuArgGlyProGlyProArgGlyGlyLeuGlyGlyProGly.. 1259

443 TGACACAGACTTCCCGCGCCACCGCGCGGAGACCTTGCACACATTTTG 492
   |||||  |||||  |||||  |||||
1260 .....ProGlyHisValGlyArg.....G 1266
```

[illegible]

304 9A19A19.


```

96 GGTGCGGACGAGC.....CCATCTGGCGGCTTCTCTCGTCGG 59
||||| |||:|||||
316 laArgAspAlaAlaProLeuArgProAlaGlyArgAlaGlnArg 332
||||| |||:|||||
58 CTGGTTCCCGCGCGGTGGCGCCACCTGGCTGGAACAGAGT 13
|||:||||| |||:|||||
333 Val.AlaAlaAlaArgGlyAlaValProAlaAlaAspProArg 347
|||:||||| |||:|||||

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-09-080-897-2

seq_documentation_block:
; Sequence 2, Application US/09080897
; Patent No. 5985574
; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Welcsh, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,897
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-080-897-2

alignment_scores:
Quality: 120.50 Length: 149
Ratio: 1.826 Gaps: 8
Percent Similarity: 44.295 Percent Identity: 33.557

alignment_block:
US-09-462-480-2/rev x US-09-080-897-2 ..

Align seg 1/1 to: US-09-080-897-2 from: 1 to: 1248

448 TTGTCATTACGGAGCTCACAGTCGTCTTCTGTCGCCAGTCGCTCG 399
||||| |||:|||||
556 LeuSerAlaAlaAlaIleThrValProProSerValProSerArg..... 570
||||| |||:|||||
398 TCGTCTTCTTACGCTCTCGCGAGCGGTGCGCGCGGACGACGCGG 349
||||| |||:|||||
571 .....AlaProValProProAlaProLeuProG 581
||||| |||:|||||
348 GCTGTGAGCGCGCGGATTGCGAACCTGCGCCCATCGTCCCGGACCCA 299
||||| |||:|||||

```

```

581 ly.....AspSerGlyThrIleIleProPro 590
|||:||||| |||:|||||
298 CCG.....GAGCGCGCCACCCGTCACCGAGGATCCGGCAACAGCC 258
||| ||| ||| |||
591 ProAlaProGlyAspSerThrProProProProProProProPr 607
||| ||| ||| |||
257 GCGGCGATCACCGAG.....GGGCAACCGGCTTTTCGATCAGCTGAGACA 212
||||| |||:||||| |||:|||||
607 oProProProLeuProGlyGlyThrAlaIleSerProProProL 624
||||| ||| ||| |||
211 TCAGCGCGGTGCGGTCAACGACCCACCTGCGCCA..... 177
||| ||| ||| |||
624 euSerGlyAspAlaThrIleProProProProLeuProGluGlyVal 640
||| ||| ||| |||
176 GTAGCGACTCGCGCGCGAGCAGGCC...GCGCGCGGTGGCGCTGA 130
||| ||| ||| |||
641 GlyIleProProSerSerLeuProGlyGlyThrAlaIleProProPr 657
||||| ||| ||| |||
129 TCCACGACGCGGATGTTTCGACAGCGGACTGTGCGCGAGCGGCCA 80
||||| ||| ||| |||
657 oProProLeuProGly.....SerAlaArgIleProProProPro 670
||||| ||| ||| |||
79 TCTGCGCGGTCTCTCTGCGGTGCGGTGCGCGCGGTGCGCGGCC 33
||| ||| ||| |||
671 ..ProProLeuProGlySerAlaGlyIleProProProProPro 685
||| ||| ||| |||

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-323-735-2

seq_documentation_block:
; Sequence 2, Application US/09323735
; Patent No. 6197932
; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Welcsh, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/323,735
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/080,897
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```



```
||||| 111111 111111
744 rOProProProPro 749

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-483-533-41

seq_documentation_block:
; Sequence 41, Application US/08483533
; Patent No. 6172047
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; APPLICANT: Chou, Joany
; TITLE OF INVENTION: Method for Treating Tumorigenic
; DISEASES
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,533
; FILING DATE: 07-MAR-95
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/419,853
; FILING DATE: 11-APR-95
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/861,233
; FILING DATE: 31-MAR-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, James P.
; REGISTRATION NUMBER: 28,491
; REFERENCE/DOCKET NUMBER: 28097/32742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-483-533-41

alignment_scores:
Quality: 118.50 Length: 222
Ratio: 1.411 Gaps: 9
Percent Similarity: 37.838 Percent Identity: 26.126

alignment_block:
US-09-462-480-2/rev x US-08-483-533-41 ..
Align seg 1/1 to: US-08-483-533-41 from: 1 to: 355

440 ACGGAGCT.....CACCAGTCGTCCTTCGTCCTCCAGTCGTCCTC 400
|||||
19 ThrGlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSe 35

399 GTCGCTCTCTTCAGCGCTCCGCGGAGCGGTGCGCGCGCGACGACCG 350
|||||
35 rGluProAlaValArgSerAlaProAlaAlaProProProProA 52
|||||
349 GGCTGGTGAGCGCGCGGATTGC..... 327

::
52 laGlyGlyProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis 68
|||||:::||||
326 .....GACCCTGGCCCATCGC 310
::: |||||::
69 ValProGluSerAlaSerAspAspAspAspTrpProAspSe 85
|||||:::|||||
309 TCCC.....GGACCCACCGGAGCGCG... 288
::: |||||:::|||||
85 rProProGluSerAlaProGluAlaArgProThrAlaAlaProA 102
287 ..CCACCCGTCACGACGATCCGGCA.....ACA 261
||||| |||||
102 rgProGlyProHisArgProAlaTrpAlaArgGlyAlaGlyLeuThr 118
260 GCCCGCGCATCACCGAGGGGCAACGGCTTTTCGATCAGCTGAGACAT 211
|||||::: |||||::: ||||| |||||
119 .ProProThrProProArgAlaProSerAlaPheArgAlaSerProS 135
210 CAGCGCGTGGGTCAACGACCCAGCTGGCGCAGGTAGGACTCCGCGC 161
|||||::: |||||
135 erAlaCysAlaSerProArgSerThrTrpArgAlaCysAlaCysAspAla 151
160 GCACGAGC..... 152
|||||
152 ArgAlaGlyArgGlyArgSerProProArgProProArgProProAr 168
151 .....CCGCGCGCGCTGGGCGCTGATCCACCGACGCGG 115
||||| |||||
168 gProProArgProProProArgGlyCysAlaSerArgProThrS 185
114 ATGTTTCGACAGCGGACTGGTGGCGAGCAGCGCCATCTGCG..... 74
::: |||||
185 erGlyCysAlaThrTrpSerGlyProArgProProAlaTrpArgAla 201
74 ..... 74

202 AlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySerGlyAlaGl 218
73 ....CGCTTCCTCGTCGCTGGGTG.....CCGCGCGCGG 41
||| |||||::: |||||
218 yTrpArgArgProArgSerSerGlyArgAlaTrpGlyProArgProV 235
40 TGCGCGCCACCTGG 27
|||||
235 alProGlyProTrp 239

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-283-471A-41

seq_documentation_block:
; Sequence 41, Application US/09283471A
; Patent No. 6340673
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; APPLICANT: Chou, Joany
; TITLE OF INVENTION: Method For Treating Tumorigenic Diseases
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/283,471A
; FILING DATE: 04-APR-1999
; CLASSIFICATION: 514
```

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/861,233
;; FILING DATE: 31-MAR-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/419,853
;; FILING DATE: 11-APR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/483,533
;; FILING DATE: 07-JUN-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Zeller, James P.
;; REGISTRATION NUMBER: 28,491
;; REFERENCE/DOCKET NUMBER: 27373/32742A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312/474-6300
;; TELEFAX: 312/474-0448
;; TELEX: 25-3856
;; INFORMATION FOR SEQ ID NO: 41:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 355 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-09-283-471A-41

alignment_scores:
Quality: 118.50 Length: 222
Ratio: 1.411 Gaps: 9
Percent Similarity: 37.838 Percent Identity: 26.126

alignment_block:

US-09-462-480-2/rev x US-09-283-471A-41 ..

Align seg 1/1 to: US-09-283-471A-41 from: 1 to: 355

440 ACGGAGCT.....CACAGTCGTCCTGTCGCCAGTCGTCCTC 400
|||||
19 ThrGlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnse 35
399 GTCCGTCCTTCACGCTCCTCGCGAGCGGTGCGCGCGCGACGACCGG 350
|||
35 rGluProAlaValArgSerAlaProAlaAlaProProProProA 52
|||
349 GCCTGGTGGAGCCGCGGATTG..... 327
|||
52 laGlyGlyProProProSerCysSerLeuLeuLeuLeuArgGlnTrpLeuHis 68
|||
326GAACCCCTGGCCCATCGC 310
|||
69 ValProGluSerAlaSerAspAspAspAspAspTrpProAspse 85
|||
309 TCCC.....GGACCCACCGAGCGCG..... 288
|||
85 rProProGluSerAlaProGluAlaArgProThrAlaAlaProA 102
|||
287 ..CCACCCGTCACGACGATCCGSCA.....ACA 261
|||||
102 rgProProGlyProHisArgProAlaTrpAlaArgGlyAlaGlyLeuThr 118
|||
260 GCCCGCGCATCACGAGGGGGGAACCGCTTTTCGATCAGCTGAGACAT 211
|||||
119 .ProProThrProProArgAlaProSerAlaPheArgAlaSerProS 135
|||||
210 CAGCGCGCTCGGTCAACGACCCACCTGCGCCAGGTAGCGACTCCCGCG 161
|||||
135 erAlaCysAlaSerProArgSerThrTrpArgAlaCysAlaCysAspAla 151
|||||
160 GCAGCAGGC..... 152
|||||
152 ArgAlaGlyArgGlyArgArgSerProProArgProProArgProProAr 168

151CCGCGCCGCGCTGGGCGCTGATCCACGACGACGG 115
|||||
168 gProProArgProProArgProProArgGlyCysAlaSerArgProThrS 185
|||
114 ATGTTTCACAGCGGACTGGTCCGACGACGACGCCCATCTGCG..... 74
:||||: |||
185 erGlyCysAlaThrTrpSerGlyProArgProProAlaTrpArgAla 201
74 74
202 AlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySerGlyAlaG1 218
73CGGCTTCCTCGTCGCGTGGGTG.....CCGCGCGCGG 41
||| |||||
218 yTrpArgArgProArgSerSerGlyArgAlaTrpGlyProArgProV 235
40 TCGCGCCCACTGG 27
|||||
235 alProGlyProTrp 239

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US91-06532-3

seq_documentation_block:
;; Sequence 3, Application PC/TUS9106532
;; GENERAL INFORMATION:
;; APPLICANT: Roizman, Bernard
;; TITLE OF INVENTION: Recombinant Herpes Simplex Viruses
;; TITLE OF INVENTION: Vaccines and Methods
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
;; ADDRESSEE: Bicknell
;; STREET: Two First National Plaza Suite 2100
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60603
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US91/06532
;; FILING DATE: 19910910
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gruber, Lewis S.
;; REGISTRATION NUMBER: 30,060
;; REFERENCE/DOCKET NUMBER: 27373/8235
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312/346-5750
;; TELEFAX: 312/984-9740
;; TELEX: 25-3856
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 355 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; PCT-US91-06532-3

alignment_scores:

Quality: 118.50 Length: 222
Ratio: 1.411 Gaps: 9

Percent Similarity: 37.838 Percent Identity: 26.126

alignment_block:

US-09-462-480-2/rev x PCT-US91-06532-3 ..

Align seg 1/1 to: PCT-US91-06532-3 from: 1 to: 355

```
440 ACGGAGCT.....CACAGTCCTCTCTTCGTCCAGTCGTCTC 400
|||||
19 ThrGlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSe 35
|||||
399 GTCGTCCTTCACGCTCCTCGCGAGCGGTGGCGCGCGCAGCAGACCG 350
|||||
35 rGluProAlaValArgSerAlaProAlaAlaProProProProA 52
|||||
349 GGTGTGTGGAGCGCGGATTGC..... 327
|||||
52 laGlyGlyProProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis 68
|||||
326 .....GACCCCTGGCCATCGC 310
|||||
69 ValProGluSerAlaSerAspAspAspAspTrpProAspSe 85
|||||
309 TCCC.....GGACCCACCGGAGCGCG.... 288
|||||
85 rProProProGluSerAlaProGluAlaArgProThrAlaAlaProA 102
|||||
287 ..CCACCCGTCACCGACGATCCGGCA.....ACA 261
|||||
102 rgProProGlyProHisArgProAlaTrpAlaArgGlyAlaGlyLeuThr 118
|||||
260 GCCGCCGCGCATCACGAGGGGCAACCGCTTTTCGATCAGCTGAGACAT 211
|||||
119 .ProProThrProProAlaProSerAlaPheArgArgAlaSerProS 135
|||||
210 CAGCGGCGTGGGTCAACGACCCAGTGGCCAGGTAGCGACTCCGCGC 161
|||||
135 erAlaCysAlaSerProArgSerThrTrpArgAlaCysAlaCysAspAla 151
|||||
160 GCAGCAGGC..... 152
|||||
152 ArgAlaGlyArgGlyArgArgSerProProArgProProArgProProAr 168
|||||
151 .....CCGCGCGCGCTGGGCTGATCCACGACGACGCGG 115
|||||
168 gProProArgProProArgProProArgGlyCysAlaSerArgProThrs 185
|||||
114 ATGTTTCGACAGCGGACTGTGGCGAGCAGGCCCATCTGCG..... 74
|||||
185 erGlyCysAlaThrTrpTrpSerGlyProArgProProAlaTrpArgAla 201
|||||
74 ..... 74
|||||
202 AlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySerGlyAlaGl 218
|||||
73 ....CGGCTTCCTCGCGCTGGGTTG.....CCGCGCGCGG 41
|||||
218 yTrpArgArgProArgArgSerSerGlyArgAlaTrpGlyProArgProV 235
|||||
40 TGCGGCCCACTGG 27
|||||
235 alProGlyProTrp 239
```

THIS PAGE BLANK (USPTO)

OM of: US-09-462-480-2 to: Pending_Patents_AA_Main:* out_format : pfs

Date: Jul 22, 2002 1:35 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame n2p_model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US09462480/runat_18072002_164419_19468/app_query.fasta_1.2850
-DB=Pending_Patents_AA_Main -OFMT=fastan -SUFFIX=rapm
-GAPOP=12.000 -GAPEXT=4.500 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000
-XGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09462480 -CGN1_1_271 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-2
Query length: 524
Database: Pending_Patents_AA_Main:*
Database sequences: 3502263
Database length: 351980561
Search time (sec): 542.380000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03265-109		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03268-114		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-658-800-109		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-659-683-114		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-680-573-109		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-680-574-114		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-729-622-109		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-730-510-114		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-341-109		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-578-114		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-024-753-109		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-025-197-114		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-596-109		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-967-114		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US091_COMB.ppt:US-09-724-685-114		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US091_COMB.ppt:US-10-084-843-114		713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-050-739-92		713.00	855.31	1.5e-39	371
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-246-191-92		713.00	855.31	1.5e-39	371
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-791-171-92		713.00	855.31	1.5e-39	371
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-804-980-92		713.00	855.31	1.5e-39	371
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US-09-943-443-92		713.00	855.31	1.5e-39	371
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-050-739-90		373.00	455.73	7.1e-17	139
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-246-191-90		373.00	455.73	7.1e-17	139
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-791-171-90		373.00	455.73	7.1e-17	139
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-804-980-90		373.00	455.73	7.1e-17	139
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US-09-943-443-90		373.00	455.73	7.1e-17	139
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-29050		155.00	186.08	0.0270	3
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-21297		148.00	175.81	0.0803	4
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US00-30037-144		147.50	178.68	0.0848	314
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US01-11988-592		147.50	178.68	0.0848	314
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US02-08278-1613		147.50	178.68	0.0848	31
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-833-245-592		147.50	178.68	0.0848	314
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-17143		146.50	175.45	0.1002	4
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US00-05989-700		145.50	179.98	0.1127	200
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US-09-925-297-700		145.50	179.98	0.1127	200
/cgn2_6/ptodata/2/paa/US095_COMB.ppt:US-09-599-287A-2		145.50	172.41	0.1182	503
/cgn2_6/ptodata/2/paa/US100_COMB.ppt:US-10-078-547-2		145.50	172.41	0.1182	503
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-33033		144.50	168.29	0.1405	7
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-21881		144.00	167.87	0.1517	7
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-31950		143.50	178.58	0.1522	7

/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US01-04098A-3487 - 141.50 161.26 0.2278
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-30739 - 140.50 164.68 0.2579
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-24973 - 140.00 167.58 0.2724
/cgn2_6/ptodata/2/paa/US095_COMB.ppt:US-09-599-287A-24 - 140.00 165.75 0.2756
/cgn2_6/ptodata/2/paa/US100_COMB.ppt:US-10-078-547-24 - 140.00 165.75 0.2756

seq_name: /cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03265-109

seq documentation block:
; Sequence 109, Application PC/TUS9903265
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US99/03265
; FILING DATE: 17-FEB-1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/024,753
; FILING DATE: 18-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9532-0023-228
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US99-03265-109

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
alignment_block:
US-09-462-480-2 x PCT-US99-03265-109 ..
Align seg 1/1 to: PCT-US99-03265-109 from: 1 to: 368
1 CTGACGAGGTGACGTGTTGTTACGACGAGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyG1 242
51 CAACCCAGCCGACGAGGAGCGCGGCACATGGCTGCTCGGCACCTC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTCTCGAACCATCCGCTGCTGGTGGTGGATCAGGCCCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

```

alignment_block:
US-09-462-480-2 x PCT-US99-03268-114
Align seg 1/1 to: PCT-US99-03268-114 from: 1 to: 368
1 CTCACGAGTGACGTCGTGTTCACCCAGGTGGCGGCACCGCGGGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyGly 242
51 CAACCCAGCCGACGAGGAAGCCGCGAGATGGGCGCTTCGGCAGCACATC 100
242 YasnProAlaAspGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATTCGCTGGCTGGTGGATCAGGCCCCACGCGGGCGG 150
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTCGCGCGGAGTCGCCTACCTGGCGCGCAGGTGGCTGTGACCCG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATGTCAGCTGATCGAAAGCCGGTGGCCCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuIleGluIysProValAlaProSerValM 309
251 TGCCGCGCGGTGTGGCGGATCGTGGTGACGGTGGCGCGCTCGCGGTG 300
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCGGGAGCGATGGGCCACGGGTTCGCAATCGCGGGGTCCACACGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGTCTGGTCGCGCGGACCGCTCGCGCAGGACGCTGAAGAAGACGACG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGlnIuArgGluGluAspSpG 359
401 AGGCACGCTGGGACGAAGAGACACTCG 429
359 loAsnAsnTrpAsnGluGluAspAspTrp 368

```

seq name: /can3 6/ntdata/2/paa/US086 COMB. pep: US-08-658-800-109

```

seq_documentation_block:
; Sequence 109, Application US/08658800
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Avenue, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,800
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C3
; TELECOMMUNICATION INFORMATION:

```

```
alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503
```


TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 368 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-658-800-109

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x US-08-658-800-109 ..

Align seg 1/1 to: US-08-658-800-109 from: 1 to: 368

```
1 CTCAGCAGGTGACGTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
|||||
51 CAACCCAGCGCAGGAGGAGCGCGCAGATGGCCCTGCTCGGCACCAAGTC 100
|||||
242 YasnProAlaAspGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGCTGGTGATCAGGCCCGCGCGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGCTGCGCGGAGTGCTACCTACCTGCGCAGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
|||||
201 CAGCCCGCTGATGTCAGCTGATGCGAAAGCCGGTGGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGGCGGCGGTGTTGCGGGATCGCTGCTGAGCGGTGGCGGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGCGATGGCGCGGTTCCGAATCCGGCGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGCTCGCGCGCAGCGCTCGCGCAGGAGCGGTGAAGAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq_name: /cgn2_6/ptodata/2/paa/US086_COMB pep: US-08-659-683-114

seq_documentation_block:
Sequence 114, Application US/08659683
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 128
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/659,683
FILING DATE: 05-JUN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 368 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-659-683-114

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-2 x US-08-659-683-114 ..

Align seg 1/1 to: US-08-659-683-114 from: 1 to: 368

```
1 CTCAGCAGGTGACGTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
|||||
51 CAACCCAGCGCAGGAGGAGCGCGCAGATGGCCCTGCTCGGCACCAAGTC 100
|||||
242 YasnProAlaAspGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGTGGTGGATCAGGCCCGCGCGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGCTGCGCGGAGTGCTACCTACCTGCGCAGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
|||||
201 CAGCCCGCTGATGTCAGCTGATGCGAAAGCCGGTGGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGGCGGCGGTGTTGCGGGATCGCTGCTGAGCGGTGGCGGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGCGATGGCGCGGTTTCGAATCCGGCGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGCTCGCGCGCAGCGCTCGCGCAGGAGCGGTGAAGAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq_name: /cgn2_6/ptodata/2/paa/US086_COMB.pep.US-08-680-573-109

seq_documentation_block:

; Sequence 109, Application US/08680573
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,573
; FILING DATE: 12-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-680-573-109

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-2 x US-08-680-573-109 ..

Align seg 1/1 to: US-08-680-573-109 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTGTTTCAGGCAGGTGGCGCGCCGCGGG 50
|||||
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
|||||
51 CAACCCAGCCGACGAGGAAGCCGCGCAGATGGCGCTGCTCGGCACCAATC 100
|||||
242 YAsnProAlaAspGluAlaAlaGlnMetGlyLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCCCGCCGCGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGTGTCGCGCGGAGTGCCTACCTGGCGCAGGTGGTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCCGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerVal 309
|||||
```

```
251 TGCCGGCGGCTGTTCCGGATCGTCCGTGACGGTGGCGCCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGCGATGGCCAGGTTTCGCAATCCGGCGGCTCCACCAGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
|||||
351 GGGTGTGTCGCGCGGCGCACGCTCGCCAGGAGCGTGAAGAGACGAGC 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
|||||
401 AGGAGCGACTGGGACGAGGAGCGACTGG 429
|||||
359 luAspAspTIPAspGluGluAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US086_COMB.pep.US-08-680-574-114

seq_documentation_block:

; Sequence 114, Application US/08680574
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 133
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,574
; FILING DATE: 12-JUL-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-680-574-114

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-2 x US-08-680-574-114 ..

Align seg 1/1 to: US-08-680-574-114 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTGTTTCAGGCAGGTGGCGCGCCGCGGG 50
|||||
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
|||||
```

```
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCGCTGCTCGCACCACTC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACATCCGCTGCTGGTATCAGCCCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGTGTCGCGCGGAGTCTACCTGGCGCAGGTGGTGGTACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCGTTGCCCTCGGTGTA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TCCCGCGCGGTGTTGCCGGATCGTGGTGACGGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGCGATGGCGAGGTTCCCAATCCGCGCGCTCCACACGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTGTCGCGCGCACCGCTCGCGCAGGAGCGTGAAGAAGACGAC 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGACGACTGGGACGAGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/us087_COMB.pap:us-08-729-622-109

```
seq_documentation_block:
; Sequence 109, Application US/08729622
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas H.
; APPLICANT: Twardzik, David R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 132
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/729,622
; FILING DATE: 11-OCT-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-729-622-109

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x US-08-729-622-109 ..
Align seg 1/1 to: US-08-729-622-109 from: 1 to: 368

1 CTGACGAGGTGACGTGCTTTTCAGCCAGGTGGCGCGCACCGGCGGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
|||||
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCGCTGCTCGCACCACTC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACATCCGCTGCTGGTATCAGCCCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGTGTCGCGCGCACCGCTCGCGCAGGAGCGTGGTGGTACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCGTTGCCCTCGGTGTA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TCCCGCGCGGTGTTGCCGGATCGTGGTGACGGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGCGATGGCGCAGGTTCCCAATCCGCGCGCTCCACACGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTGTCGCGCGCACCGCTCGCGCAGGAGCGTGAAGAAGACGAC 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGACGACTGGGACGAGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/us087_COMB.pap:us-08-730-510-114

```
seq_documentation_block:
; Sequence 114, Application US/08730510
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 137
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
```

seq_name: /cgn2_6/ptodata/2/paa/US089_COMB.pep:US-08-942-341-109

seq_documentation_block:
; Sequence 109, Application US/08942341
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodges, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,341
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-730-510-114

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x US-08-730-510-114 ..
Align seg 1/1 to: US-08-730-510-114 from: 1 to: 368

1 CTGACGAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCAGCAGGAGCGGCGCAGATGGGCTGCTCGGCACCATC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACATCCGCTGCTGATCAGCCAGCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTGCGCGCGGAGTCTGCTACCTGGCGCAGGTGGTGGTTCACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATCTCTAGCTGATCGAAACCGCGGTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGCGCGCTGTTGCCGATCGTCGGTGACGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGAGCGATGGGCGCAGGGTTCGCAATCCGCGCGCTCCACACGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGGTCCGCGCGGCGCAGCGCTCGGCAGGAGCGGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAsp 359
401 AGGACGACTGGCAGCAAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x US-08-942-341-109 ..
Align seg 1/1 to: US-08-942-341-109 from: 1 to: 368

1 CTGACGAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCAGCAGGAGCGGCGCAGATGGGCTGCTCGGCACCATC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACATCCGCTGCTGATCAGCCAGCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTGCGCGCGGAGTCTGCTACCTGGCGCAGGTGGTGGTTCACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292

```
201 CACCCGCTGATGCTCAGCTGATCGAAGAACCGGTTGCCCGCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGGCGGGCTGTGCGGATCGTGGTACGGGTGCGCGGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGCGATGGCCAGGTTCCCAATCCGGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGGTCCGCGCGCACCGTCCGCGCAGGAGCGTGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspG 359
|||||
401 AGGACGACTGGGACGAAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US089_COMB.pep:US-08-942-578-114

```
seq_documentation_block:
; Sequence 114, Application US/08942578
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/942,578
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-942-578-114
```

alignment_scores: Quality: 713.00 Length: 143
 Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

```
alignment_block:
US-09-462-480-2 x US-08-942-578-114 ..
Align seg 1/1 to: US-08-942-578-114 from: 1 to: 368

1 CTCAGCAGGTGACGCTGCTGTTGTCAGCCAGGTGGCGCGCACCGCGCGCG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
|||||
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGGCTCTCGGCACCAAGTC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCGAACCACCTCCGCTGGTGGATCAGGCCCGCCAGCGCGGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGTGTCGCGCGGAGTCGCTACCTGCGCGCAGGTGGGTGGTTCGAC 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CAGCCGCTGATGTCAGCTGATCGAAGAACCGGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGGCGGGCTGTGTTGCCGGATCGTGGTACGGGTGCGCGGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGCGATGGCCAGGTTCCCAATCCGGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGGTCCGCGCGCACCGTCCGCGCAGGAGCGTGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359
|||||
401 AGGACGACTGGGACGAAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-024-753-109

```
seq_documentation_block:
; Sequence 109, Application US/09024753
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 236
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/024,753
; FILING DATE: 18-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
```

TUBERCULOSIS

```
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-024-753-109

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
  Percent Similarity: 98.601      Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x US-09-024-753-109 ..
Align seg 1/1 to: US-09-024-753-109 from: 1 to: 368

1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242

51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCGCTGCTCGGCACCATC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTGCAACATCCGCTGGTGGTATCAGGCCCGCAGCGCGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCTGTGTCGGCGGAGTCGCTACCTGGCGCAGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292

201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCGCTTCCCGCTCGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerVal 309

251 TGCCGGCGGTGTTCGGGATCGTGGTGGTGGTGGTGGTGGTGGTGG 300
|||||
309 etProAlaAlaAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThr 342

301 GGTCTGGTCCGCGCGGACCGCTCGCGCAGGACGCTGAAGAACGAC 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAsp 359

401 AGGACGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-025-197-114
seq_documentation_block:
; Sequence 114, Application US/09025197
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
```

```
351 GGGTCTGCTCGCGCCGACCGCTCGCGCAGGAGCGTGAAGAACGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGACGACTGGGACGAAGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluAspAspTrp 368
```

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-072-596-109

```
seq_documentation_block:
; Sequence 109, Application US/09072596
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-109
```

alignment_scores: Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block: US-09-462-480-2 x US-09-072-596-109 ..

Align seg 1/1 to: US-09-072-596-109 from: 1 to: 368

```
1 CTGACGAGGTGACGTGCTGTTTCAGCAGGTGGCGGCACCGCGCGGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
|||||
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGCGCTCTCGGCACCACTC 100
|||||
```

```
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCTGAACCATCCGCTGGTGGATCAGGCCCGCCAGCGCGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTCTGCTGCGCGGAGTCTACCTACCTGGCGCAGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CAGCCCGCTGATGTCTCAGCTGATCGAAAGCCCGTTGCCCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGC CGCGCGCTGTTCGCGGATCGTCCGTCGAGCGGTGGCGGCTCCCGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGCGGAGCGATGGCGCAGGGTTCCGAATCCGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGCTGCGCGCGCGCACCGCTCGCGCAGGAGCGTGAAGAACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGACGACTGGGACGAAGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluAspAspTrp 368
```

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-072-967-114

```
seq_documentation_block:
; Sequence 114, Application US/09072967
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
```

```
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-114

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x US-09-072-967-114 ..
Align seg 1/1 to: US-09-072-967-114 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGTTCTCAGCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242

51 CAACCCAGCCGACGAGGAGCGCGCAGATGGGCTGCTCGGCACCGTC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTGAACCATCCGCTGGCTGGTGGATCAGGCCCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275

151 GGCCTGTGGCGGAGTACCTACCTGGCGCAGTGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292

201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCGGTTGCCCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValm 309

251 TGCCGGCGGCTGTTCGGGATGTCGGTGACGGTGGCGCGCGCTCGG 300
|||||
309 etProAlaAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 325

301 GGTCCGGGAGCGATGGCGAGGTTCGCAATCCGCGCGCTCCACGCCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342

351 GGGTCTGTGTCGCGCGGCGCACCGCTCGGCGAGGAGCGGTGAAGAAGCAGCAG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359

401 AGGACGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /cqn2_6/ptodata/2/paa/US097_COMB.pep:US-09-724-685-114

seq_documentation_block:
; Sequence 114, Application US/09724685
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
;             Skeiky, Yasir
;             Dillon, Davin C.
;             Campos-Neto, Antonio
; TITLE OF INVENTION: Compounds and Methods for
;                   Immunotherapy and Diagnosis of Tuberculosis
; NUMBER OF SEQUENCES: 155
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```



```

309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCGGGAGCGATGGGCCAGGTTCCGAATCGGGCGGCTCCACCGAGCCC 350
   ||| ||||| ||||| ||||| ::||| ||||| ||||| |||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGCTCTGGTCGCCCGCCACCGCTCGCGCAGGAGCGGTGAAGAAGACGACG 400
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGACGACTGGGACGAGAGAGGACTGG 429
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
359 luAspAspTrpAspGluGluAspTrp 368

```

THIS PAGE BLANK (USPTO)

seq_documentation_block:

```
; Sequence 252, Application US/09994404
; GENERAL INFORMATION:
; APPLICANT: ESSER, KLAUS M.
; APPLICANT: CHAN, JOHN Y.
; APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN
; APPLICANT: DELVECCIO, ALFRED MICHAEL
; APPLICANT: DILLON, SUSAN B.
; APPLICANT: LEARY, JEFFREY JOSEPH
; APPLICANT: SUTTON, DAVID
; TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES
; TITLE OF INVENTION: SIMPLEX VIRUS TYPE-2
; FILE REFERENCE: P50583
; CURRENT APPLICATION NUMBER: US/09/994,404
; CURRENT FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 09/297,477
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/030,279
; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: US 60/049,018
; PRIOR FILING DATE: 1997-06-09
; NUMBER OF SEQ ID NOS: 303
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 252
; LENGTH: 3051
; TYPE: PROT
; ORGANISM: Herpes simplex
US-09-994-404-252
```

alignment_scores:

```
Quality: 133.00 Length: 253
Ratio: 1.357 Gaps: 13
Percent Similarity: 38.735 Percent Identity: 26.877
```

alignment_block:

```
US-09-462-480-2/rev x US-09-994-404-252 ..
```

```
Align seg 1/1 to: US-09-994-404-252 from: 1 to: 3051
```

```
502 CTTCTCTCGCAAAATGTCGAAGTCTCCGGCCGGGTGGCGGGAAG 453
||||| ||| : : : : : |||||
2582 ProSerProAlaAspAlaAlaLeuProProAlaPheSerGlyse 2598
| : : : : : |||
452 TCTGTCCTCATACGGAGCTCACCACTGCTCTTCGTCGCCAGTCGC 403
| : : : : : |||
2598 rAlaAlaAlaPheSerAlaAlaValProArgValArgSerArgT 2615
| : : : : : |||
402 CTCGTCCTCTCTCAGCTCTCTCGCGGAGGGTGCGG..... 365
||| :||| ||| : : : : |||
2615 hrArgAlaLysSerArgAlaProArgAlaSerAlaProProGluGlyTrp 2631
||| :||| ||| : : : : |||
364 .....GC GCG 360
2632 ArgProProAlaLeuProAlaProValAlaProValAlaAlaSerAlaAr 2648
359 ACCA.....GACCCGGGCTGGTGA 340
||| :||| ||| : : : : |||
2648 gProProAspGlnProProThrProGluSerAlaProProAlaTrpValS 2665
339 GCCGCCGGATGGCAACCTGGCCCATCG..... 311
||| :||| ||| : : : : |||
2665 erAlaLeuProLeuProProGlyProAlaSerArgAlaPheProAlaPro 2681
310 .....CTCCGGACCCACCG..... 296
2682 ThrLeuAlaProIleProProProProAlaGluGlyAlaValAlaProGI 2698
295 ....GAGCGGGCCAC.....CCGTCACCGA 274
: : : : : |||||
2698 yAspAspArgArgGlyArgArgGlnThrThrAlaGlyProSerProT 2715
273 CGATCCGGCAACACCCCGCGCA..... 251
```

```
|| : : : ||| |||
2715 hrProProArgGlyProAlaAlaGlyProProArgLeuTrpAlaVal 2731
250 .....TCACCGAGGGGCAACC 234
2732 AlaSerLeuSerAlaSerLeuAsnSerLeuProSerProArg...AspPr 2747
233 GGCTTTTCGATCAGTCAGACATCAGCGGCGTGGGGTCAACGACCCACC 184
||| : : : : : ||||| : : : : |||
2747 oAlaAspHisAlaAlaValSerAlaAlaAlaAlaVal...ProPr 2763
183 TGCGCCAGGT.....AGCGACTCCGCGCAGCAGCAGCCCGCGCCGC 143
| : : : : ||| : : : : |||
2763 oSerProGlyLeuAlaProProThrSerAlaValGlnThrSerProProp 2780
142 CGCTGGGGCTGATCCA.....CCAGCCAGCGGATGGTTC 108
||| : : : ||| ||| |||||
2780 roLeuAlaProGlyProValAlaProSerGluProLeuLeuCysGlyTrp... 2795
107 GACAGCGGACTGTGTCGAGCAGCGCCATCTGCGCGGCTTCCTCGTCGCG 58
: : : : : ||||| : : : : |||
2796 .....ValValPro.GlyGlyProValAlaArgPro..... 2806
57 TGGTTGCCGCGCGGTGCGCCGCTGCTGAACAACAGCGTCACCT 8
||||| : : : : : |||
2807 .....ProProGlnSerProAlaThrLysPro 2815
7 GCTGCA 2
|||||
2816 AlaAla 2817
```

```
seq_name: /cgn2_6/ptodata/2/paa/us10_NEW_COMB.pep:US-10-155-881-22036
```

seq_documentation_block:

```
; Sequence 22036, Application US/10155881
```

```
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 22036
; LENGTH: 261
; TYPE: PROT
; ORGANISM: Glycine max
US-10-155-881-22036
```

alignment_scores:

```
Quality: 127.50 Length: 181
Ratio: 1.401 Gaps: 11
Percent Similarity: 50.276 Percent Identity: 35.359
```

alignment_block:

```
US-09-462-480-2/rev x US-10-155-881-22036 ..
```

```
Align seg 1/1 to: US-10-155-881-22036 from: 1 to: 261
```

```
476 CTTCCGGCCGGGTGGCGGAGTCTGTTCATTACGGGAGCTCACCA 427
||||| : : : : : ||||| : : : : |||
65 LeuProArgThrAlaThrAlaArgSerThrAlaAlaGlyAlaAlaPheAl 81
426 GTCGTCCTCTTCGTCGCCAGTGTCTGCTCTTCTTTCAGCTCCTCGC 377
||||| : : : : : ||||| : : : : |||
81 aCysArgSerSerAlaProAlaSerSerSerSerArgAlaSerSera 98
376 CGAGCGGTGCGG...GCGCGACCAACACCCGGGTGGTGGAGCCCGCGGAT 330
```

```
98 laThrSer.ProThrAlaArgProSerSerGlySerSerAlaArgArg.. 113
|||||
329 TGCACCAACCGCCCATCG...CTCCGGAGCCACCGCGGCGCCACC 283
|||||
114 ....AsnProProSerSerProLeuProAlaProAlaProLeuArgArgA 129
|||||
282 CGTCACGACGATCCGCAACAGCCGCGGCATCACCAGGGGCAACCG 233
|||||
129 laSerProLeu.....SerProThrThrSerProSerSerLeuPro 142
|||||
232 GCTTTTCATCAGTCAGATCAGCGCGGTGCGGGTCACGACCCACCT 183
|||||
143 LeuProProSerSerSerAlaAsnAlaPheAlaLeuThrThrThr..Pro 158
|||||
182 CGCGCAGGTACGACTCGCGCGCAGCAGGCC.....GCGCCGC 142
|||||
159 ProProLysThrThrProSerArgSerSerProProProArgArgProPr 175
|||||
141 GCTGGGCGCTGATCCACGACGCGGATGGTTCGACAGCGGACTGGTGC 92
|||||
175 oSerGlyProTyrLeuProAlaProThrSerAlaArgSerGlyAlaSerP 192
|||||
91 CGACGAGGCC...ATTCGCGCGCTTCCTCGTGGGTGGTTC..... 50
|||||
192 roProArgProAsnSerCys.ArgCysArgArgArgilleGlnCysPheTh 208
|||||
49 .....CGCGCGGTGCGGCCACCT..... 29
|||||
208 rIleThrIleAsnThrThrSerSerAsnSerAsnArgProProTrpGluA 225
|||||
225 rgProLeuProGlnGlySerGlyIleThrSerProAla 237
|||||
```

seq_name: /cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep.US-10-155-881-25973

```
seq_documentation_block:
; Sequence 25973, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 25973
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-155-881-25973
```

```
alignment_scores:
Quality: 126.50 Length: 199
Ratio: 1.360 Gaps: 12
Percent Similarity: 46.734 Percent Identity: 30.653

alignment_block:
US-09-462-480-2 x US-10-155-881-25973 ..
Align seg 1/1 to: US-10-155-881-25973 from: 1 to: 318

12 GACGTCTGTTGTTCAGCGGCGGCGCACCGCGCGGCGCAACCCAGCCG 61
|||||
120 AspAlaPheProArgAlaGlyAspArgGlnArgArgHis.....Ar 134
|||||
62 ACGAGGAAGCGCGCAGATGGCGCTGCTCGGCACCGTCCGCTGTGCAAC 111
```

```
134 gArgGly.....HisArgGlyLeuLeuPro 143
|||||
112 CATCCGTGGCTGGTATCAGCCCGCCAGCGCGGCGCGC..... 149
|||||
143 roGlnAlaAlaArgLeuArgAlaGlnArgGluArgArgArgHisGln 159
|||||
150 .....GGGCCTGCTGCGCGGAGTACCTACCTGGCGCAGG..... 185
|||||
160 ArgHisAlaProAlaAlaArgGly.....AspTrpGlyArgArgArgAr 174
|||||
186 .....TGGTCTGTTACCCG..... 200
|||||
174 gAlaAlaGlyGluValArgAspIleValHisValTrpArgLeuProG 191
|||||
201 .....CACGCCGCTGATGCTCAGCTGATCGCAAA 230
|||||
191 lyAlaGlyAlaAlaArgGlyHisGlyAlaArgValProArgArgArg 207
|||||
231 GCCGTTGCCCTCGGTGATGCGCGGCGCTGTTGCCGATCGTCGGTGA 280
|||||
208 AlaGly.....AlaGlyGly..... 212
|||||
281 CGGTGCGCGCGCTCCGCTGGGTCCGGAGCGATGGCCAGGGTTCGCAA 330
|||||
213 .GlyTrpGluArgHisGlyGlyAlaAspArgValGlyProArgHisGly 229
|||||
331 TCCCG.....CGGCTCCACAGCCCGGCTGCTGCGCGCGCGCAC 371
|||||
229 spArgGlyHisValArgGlnArgHisValArgGluAlaAlaAlaGlyPro 245
|||||
372 GCTCGCGCAGGCGGTGAAGAAGACGACGAGCGACTGGGACGAAGAGG 421
|||||
246 GlyArgGlyGly.....ArgArgAlaValArgValGlyGlyArgAr 260
|||||
422 ACGA.....CTGTTAGCTCCCGTAATGACAACAGACTTCCCGGCC 462
|||||
260 gArgAlaAspGlyAlaAlaGluGlnArgArgArgArgAlaProAla 277
|||||
463 ACCCGCGCGGAAGACTTGCACACATTTTGGCGGAGGAGTAAG 507
|||||
277 spValArgArg.ArgProAlaAspAlaAlaAlaArgHisValArg 291
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.US-60-382-898-278

```
seq_documentation_block:
; Sequence 278, Application US/60382898
; GENERAL INFORMATION:
; APPLICANT: Hudson, Keith
; APPLICANT: et al.
; TITLE OF INVENTION: Plant Receptors and Ligands
; FILE REFERENCE: 1066P
; CURRENT APPLICATION NUMBER: US/60/382,898
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 1344
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 278
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-60-382-898-278
```

```
alignment_scores:
Quality: 125.50 Length: 175
Ratio: 1.793 Gaps: 7
Percent Similarity: 40.000 Percent Identity: 29.143

alignment_block:
US-09-462-480-2/rev x US-60-382-898-278 ..
Align seg 1/1 to: US-60-382-898-278 from: 1 to: 708
```

```
421 CCTCTCTGTCACGTCGTCCTGCTGCTTCTTCACGCTCCTGCGCAGC 372
||||| ||||| ||| |||||
38 ProLeuProProSerAlaPro.....ProProAsnArgAl 49
||||| ||||| |||||
371 GGTGCGCGCGGACGACGACCGCGGCTGTGGAGCGCGCGGATTGCGAAC. 323
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49 aProProProProProProValThrSerProProProValAlaAsnG 66
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
322 .....CCTGGCCATCGCTCCGGACCGCGGAGCGCGGCAC..... 284
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 lAlaProProProProLeuProLysProProGluSerSerSerProPro 82
CCGTCACGACGATCCGACACAGCCGCGCA...TCACGAGGGGCA 237
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
83 ProGlnProValIleProSerProProSerThrSerProProProGln 99
ACCGGCTTTTCAGTCAGTCAGACATCAGCGGTG..... 201
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 nProValIleProSerProProProSerAlaSerProProProAlaLeuV 116
201 ..... 201
116 alProProLeuProSerProProProProAlaSerValProProPro 132
201 ..... 201
133 ArgProSerProSerProProIleLeuValArgSerProProProSerVa 149
200 .....CGGTCAAGACCCACCTGCGCCAGGTAGCGACTCCGCGCGCACA 155
149 largProIleGlnSerProProProProProProSerAspArgProThrGlnS 166
154 GCGCGCGCGCGCTGGCGCTGATCCACGACGCGCGGATGTTTCGAC 105
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
166 erProProProProSerProProSerProProSerGluArgProThrGln 182
104 ACGGACTGTGCGGACGAGCGCCATCTGCGCGCTTCTCTGCGGCTGG 55
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
183 Ser.....ProProProProSerGluArgProThrGlnSerPro... 196
54 GTTCCGCGCGCGGTGCGCGCCACC 30
197 ....ProProProProSer 203
```

seq_name: /cgn2_5/ptodata/2/paa/US09_NEW_COMB.pep:US-09-667-170A-187

seq_documentation_block:

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Henderson, Robert A.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Fling, Steven P.

; APPLICANT: Mohamath, Raodoh

; APPLICANT: Algate, Paul A.

; APPLICANT: Secrist, Heather

; APPLICANT: Indirias, Carol Yoseph

; APPLICANT: Benson, Darin R.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

; THE THERAPY AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.475C7

; CURRENT APPLICATION NUMBER: US/09/667,170A

; CURRENT FILING DATE: 2002-05-03

; NUMBER OF SEQ ID NOS: 440

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 187

; LENGTH: 595

; TYPE: PRT

; ORGANISM: Homo sapien

US-09-667-170A-187

alignment_scores:

..

Quality: 123.50 Length: 170
Ratio: 1.816 Gaps: 8
Percent Similarity: 40.000 Percent Identity: 28.824

alignment_block:

US-09-462-480-2/rev x US-09-667-170A-187 ..

Align seg 1/1 to: US-09-667-170A-187 from: 1 to: 595

```
490 AAATGTTGCGAAGTCTCCGG.....CCCGGGTGGCGGGGAAGTCTGTT 447
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
474 ArgLeuTrpArgGlnPheargvalglyglypnePro..... 486
446 GTCATTACGGGAGCTCACGAGTCGTCCTTCTGTCCTCCAGTCGTCCTCGTC 397
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
487 .....ProProProProSerArgProProA 495
396 GTCCTCTTCA.....CGCTCTCTGCGGAGGCGGTCCGCGGG 360
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
495 laValLeuLeuProLeuLeuArgLeuAlaCysAlaGlyAspProGlyAla 511
359 ACCAGACCCGGCTGGTGAGCGCGCGGATTGCGAACCTGCGCCCATCGC 310
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
512 ThrArgProGlyProArgArgPro.....Al 520
309 TCCCGGACCCACCGGAGCGCGCCACCGTCACCGACGATCGGCAACA. 261
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
520 aArgArgProArgGlyGluLeuIleProArgProArgProAlaAlap 537
260 .....GCCGCGGCATCACCGAGGGGCAACC 234
537 roSerGluGluGlyLeuArgMetGluSerSerValAspGlyAlaThr 553
233 GGCCTTTTCGATCAGTCAGACATCAGCGCGGTGCGGGTCAACGACCCACC 184
553 ..... 553
183 TCGCGCAGGTAGCGACTCCGCGCGCAGCAGCGCGCGCGCGGTGGGCG 134
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
554 .AlaThrThrAlaAspAlaAla...SerGlyGluAlaProGluAlaGlyP 569
133 CTGATCCACGACGCGGATGTTGCGACGCGGACTGTCGCCGAGCAGG 84
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
569 roSerProSerHisSer..... 574
83 CCCATCTCGCGGCTTCTCTGTCGCTGCGGTGCGCGCGCGGTCCGCC 34
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
575 ProThrMetCysGlnThrGlyGlyProGlyProProProGlnProPr 591
33 CACCTGGCTG 24
||| |||||
591 oArgTriLeu 594
```

seq_name: /cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-452

seq_documentation_block:

; Sequence 452, Application US/10080170B

; GENERAL INFORMATION:

; APPLICANT: COLE, S.T.

; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR

; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PHYLOGENESIS OR

; TREATMENT OF MYCOBACTERIOSES

; FILE REFERENCE: 03495.0218

; CURRENT APPLICATION NUMBER: US/10/080,170B

; CURRENT FILING DATE: 2002-06-10

; PRIOR APPLICATION NUMBER: 60/270.123

; PRIOR FILING DATE: 2001-02-22

; NUMBER OF SEQ ID NOS: 652

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 452

; LENGTH: 227

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

US-10-080-170B-452

alignment_scores:
Quality: 123.00 Length: 220
Ratio: 1.281 Gaps: 11
Percent Similarity: 43.636 Percent Identity: 26.364

alignment_block:
US-09-462-480-2/rev x US-10-080-170B-452 ..

Align seg 1/1 to: US-10-080-170B-452 from: 1 to: 227

464 GTGGCCGGGAAGTCTGTGTCTATACGGGA.....GCTCAC..... 429
|||||.....
15 ValGlySerSerAlaAlaLeuLeuThrGlyGlyIleAlaHisAlaAspPr 31
428CAGTCGTCTCTTCGT 413
31 oAlaProAlaProAlaProAlaProAsnIleProGlnGlnLeuIleSerS 48
412 CCCAGTCGTCTCGTCGCTTCTTCCAGCTCTCTGCGCAGCGGTCCCGCC 363
|||.....
48 erAlaAlaAsnAlaProGlnIleLeuGlnAsnLeuAlaAlaLeuGly 64
362 GCGACACAGCCCGGCTGGTGGAGCCCGGATTGCGAACCCCTGCCCAT 313
||||| ||| |||
65 Alathr....ProProLeuSerAlaProLysValAlaGluProAlaProAl 80
312 CGTCCCGGACCCAGCGGAGCGGCCACCGTCACCGAC...GATCCGG 266
||||| |||
80 aAlaProGlyIleThrAlaThrPheProGlyLeuThrProAlaAlaProA 97
265 CAACAGCCGCC..... 255
||||| |||
97 laAlaAlaAlaAlaProAlaLeuThrProSerIleProGlyValAsnAla 113
254GGCATCACCGAGGGGCAACCGCTTT..... 228
114 ProfileProGlyIleThrProAlaAlaProAlaLeuProValThrAlaPr 130
227TCGATCAGCTGAGACATCAGCGCGCTGC 200
130 oAlaAlaAlaProThrIleProGlyValAsnAlaProIleProGlyIleT 147
199 GGGTCAACGACCCACTCGG.....CCAGGTAGC 171
147 hrAlaProAlaProAlaAlaAlaValProAlaSerValProGlyVal 163
170 GACTCCGGCGCG..... 159
164 ProSerAlaLysValAspLeuProGlnLeuProTyrLeuProLeuGlnVa 180
158ACGAGCCCGCG.....CCGCGCTGGGCGCTGATC 128
180 lProGlnGlnLeuSerLeuProAlaAspLeuProAlaLeuAla..... 194
127 CACACGACGAGGATGGTTCGACAGCGGACGTGGTCCGCGAGGACGCCATC 78
195SerGlyValIleProAlaAlaProIle 203
77 TCGCGGGCTTCTCTCGCTGGCTGGGTGGCG.....CCGCGGTGCCGCC 34
204 AlaProThrProAlaProGlyAlaProAlaLeuProProGlyProPr 220
33 CACTGGGTG 24
220 oSerLeuLeu 223

seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10704

seq_documentation_block:
; Sequence 10704, Application US/10155881

GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 10704
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-10704

alignment_scores:
Quality: 122.00 Length: 154
Ratio: 1.649 Gaps: 9
Percent Similarity: 48.052 Percent Identity: 34.416

alignment_block:
US-09-462-480-2/rev x US-10-155-881-10704 ..

Align seg 1/1 to: US-10-155-881-10704 from: 1 to: 356

502 CTTCTCTCCCAAAATGTTGCCAGTCTTCGCGCCGCGGTGGCCGGAAG 453
||||| |||
207 ProSerSerSerThrThrThrThrThrProAlaAlaPro..... 221
452 TCTGTTGTCATTACGGGAGCTCACC....AGTCGTCTCTTCTGTCACGTC 406
||| ||| |||
222LeuProCysArgArgCysThrProProThrProThrArgSerThrA 237
405 GTCCTCTCGCTCTTCTCACGCTCTCGCGCA.....GCGGTGCGC 365
||| |||
237 rgGlnArgArgArgGlnArgThrProAlaArgThrAlaAlaAlaSerPro 253
364 GCGGACACGACCGCGGCTGTGGAGCCGCGGATTCGGAACCCGCGGCC 315
: ||| |||
254 ThrArgProPro.....ArgAlaProProThrSerPr 265
314 ATCGTCCCGGACCCACCGGAGCGG.....C 289
||||| |||
265 oSerLeuProProProProAlaSerThrArgValAlaThrThrTrpProA 282
288 GCCACCCGTCACCGACATCCGGCAACAGCCGCGCATCACCGAGGGG 239
||| |||
282 rgAlaProArgArgSerHisArgAlaArgAlaProProSerProArgArg 298
238 CAACGCGCTTTTCGATCAGCTGAGACATCAGCGGCGTGGGTCAACGAC 189
: ||| |||
299 ArgArgArg...ArgAlaProThrThrArgSerArgCysProSerSerTh 314
188 CCACCTGCGCCAGGTAGGACTCCCGCGCAGCAGGCGCG.....CGCCGC 142
|
314 rSerSerAlaSerAlaArgArgAspGlyGlyAlaGlyProGlyArgProG 331
141 GCTGGGCTGTATCCACGACGACGCGGATGGTTCCACAGCGGACTGGTGC 92
||| |||
331 lyGlyGlyVal.....GlnGlyGlyGlyGlyGlnArgThrArgAl 345
91 CGAGCAGGCC 82
||| |||
345 aProAlaAla 348

seq_name: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:PCT-US02-09921-1022

seq_documentation_block:
; Sequence 1022, Application PC/TUS0209921

: GENERAL INFORMATION:

```

/ APPLICANT: INCYTE GENOMICS, INC.
/ APPLICANT: DAFFO, Abel
/ APPLICANT: JONES, Anissa L.
/ APPLICANT: TRAN, Alanna-Phung B.
/ APPLICANT: DAHL, Christopher R.
/ APPLICANT: GIETZEN, Barryl
/ APPLICANT: CHINN, Joyce
/ APPLICANT: DUFOUR, Gerard E.
/ APPLICANT: HILLMAN, Jennifer L.
/ APPLICANT: YU, Jimmy Y.
/ APPLICANT: TUASON, Olivia
/ APPLICANT: YAP, Pierre E.
/ APPLICANT: AMSHEY, Stefan R.
/ APPLICANT: DAUGHERTY, Sean C.
/ APPLICANT: DAM, Tam C.
/ APPLICANT: LIU, Tommy F.
/ APPLICANT: NGUYEN, Duy-Viet An
/ APPLICANT: KLEEFELD, Yael
/ APPLICANT: GERSTIN JR., Edward H.
/ APPLICANT: PERALTA, Careyna H.
/ APPLICANT: DAVID, Marie H.
/ APPLICANT: LEWIS, Samantha A.
/ APPLICANT: CHEN, Alice J.
/ APPLICANT: PANZER, Scott R.
/ APPLICANT: HARRIS, Bernard
/ APPLICANT: FLORES, Vincent
/ APPLICANT: MARWAHA, Rakesh
/ APPLICANT: LO, Audrey
/ APPLICANT: LAN, Ruth Y.
/ APPLICANT: URASHKA, Michael
/ TITLE OF INVENTION: SECRETORY MOLECULES
/ FILE REFERENCE: PT-1232 PCT
/ CURRENT APPLICATION NUMBER: PCT/US02/09921
/ CURRENT FILING DATE: 2002-03-27
/ PRIOR APPLICATION NUMBER: 60/280, 067; 60/280, 068; 60/291,280
/ 60/291, 829; 60/299, 428; 60/300, 001; 60/299, 776
/ PRIOR FILING DATE: 2001-03-29; 2001-03-29; 2001-05-16; 2001-0
/ 2001-03-17; 2001-06-19; 2001-06-20; 2001-06-20
/ NUMBER OF SEQ ID NOS: 1146
/ SOFTWARE: PERL Program
/ SEQ ID NO 1022
/ LENGTH: 380
/ TYPE: PPT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ OTHER INFORMATION: Incyte ID No: LG:007916.8.orf2:2001JUN22
/ PCT-US02-09921-1022

```

alignment_scores:		
Quality:	121.50	Length: 167
Ratio:	1.293	Gaps: 10
Percent Similarity:	56.287	Percent Identity: 34.132

alignment_block:
US-09-462-480-2/rev x PCT-US02-09921-1022

Align seq 1/1 to: PCT-US02-09921-1022 from: 1 to: 380

```

479  AGTCTCCGCGCCGGGTGGCCGGGAAGTCTGTTGTCTATTACGGGAGCTCA 430
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
137  SerLeuTyrAlaAspSerAspGlyGluAsnAspSerThrSerAspSerG1 153
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429  CCAAGTCGTCTCTTCGTCCAGTCGTCTCTGCTGCTTCTTCACGCTCCT 380
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153  uSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerS 170
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
379  GC CGC GCG CGT GTC GCG CGC GCG CAC CAG CCG GGG CTG GTG GAG CCG CGG GAT 330
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170  erSerSerSerSerSerSerSerGluSerSerSerGluAspGluGlu 186
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

329 TGCGAACCTGGCCCATGCTCCGG.....GACCCACCGAGCGCGCGC 286
||| ::::| | | | | | | | | | | | | | | | | | | | | | | | | | |
187 GluGlu.ArgProAlaLaLeuProSerAlaSerProProArgGluV 203

285 ACCCGTCACCGAGC.....ATCCGCCAACCAGCGC 257
||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
203 alProValProThrProAlaProValGluValProGluArgVal 219

256 CCGGCATCACCGAGGGCAACCGCTTTTCGATCAGCTGAGACATCAGC 207
::::| | | | | | | | | | | | | | | | | | | | | | | | | | |
220 AlaGlySerProValThrProProAlaArgThrGlyGlyValSerSerLy 236

206 GGCGTGC.....GGGTCAACAGCACCCACTGGCCGAGTAGCGACTC 166
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
236 sAlaCysArgProHisGlyGlyValThrProGlnCysAlaSerAlaSerP 253

165 CGCGCGCAGCA.....GGCCCGCGCCCGCTGGGGCTGATCCACCAG 122
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
253 roArgTrThrCysTrpAlaProArgPro.....LeuProHisAla 266

121 CCAGCGGATGTTTCACACGCGACTGGTGCCGAGCAGGCGCCATCTCGCGC 72
||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
267 ProMet.....SerValPro..LeuLeuProSer...ProSerCysPro 279

71 GCTTCCTCGTGGCTGGG...TTCCGCGCCGCGGTGGCGCCCACTGG 27
:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
280 HisProArgAsnAlaGlyLvsLeuSerProSerLeuProSerAlaTrp 295

seq_documentation_block:

```

? Sequence 3962, Application US/60360039
?
? GENERAL INFORMATION:
?
? APPLICANT: Cao, Yongwei
? APPLICANT: Chen, Xianfeng
? APPLICANT: Goldman, Barry S.
? APPLICANT: Hinkle, Gregory J.
? APPLICANT: Slater, Steven C.
?
? TITLE OF INVENTION: EXPRESSION OF MICROBI
?
? FILE REFERENCE: 38-10,52052A
?
? CURRENT APPLICATION NUMBER: US/60/360,039
?
? CURRENT FILING DATE: 2002-02-21
?
? NUMBER OF SEQ ID NOS: 47374
?
? SEQ ID NO 3962
?
? LENGTH: 639
?
? TYPE: PRT
?
? ORGANISM: Neurospora crassa
?
? US-60-360-039-3962

```

alignment_scores:		
Quality:	121.50	Length: 189
Ratio:	1.538	Gaps: 12
Percent Similarity:	41.799	Percent Identity: 31.217

alignment_block:

US-09-462-480-2/rev x US-60-360-039-3962

Align seq 1/1 to: US-60-360-039-3962 from: 1 to: 639

473 CCGGCCGGGTGCGCCGAAGTCTGTTGTCAATTACGGGAGTCACCAGTC 424
||||| :||| :
319 ProAlaArgSerGlyLys.....LeuAspThrGluAsnHisGln.G 333
:
423 GTCTCTTCGTCCCAGTCGT..... 404
:||| :
333 LuProAlaProProArgPheAlaValProProIleAlaAspAla 349

403CCTCGTGCTCTTCACGCCTCCTGC 378
||||| :
350 GlyLysPheAlaHisSerAspProProArqHisThrProSerAlaProGI 366


```
alignment_scores:
  Quality: 119.50      Length: 135
  Ratio: 1.757         Gaps: 4
  Percent Similarity: 50.370  Percent Identity: 28.889

alignment_block:
US-09-462-480-2/rev x US-09-573-655B-274 ..
Align seg 1/1 to: US-09-573-655B-274 from: 1 to: 312

425 TCCTCTCTTCGTCGCCAGTCCTCGTCCTCTTCTTACGCTCCTGCGC 376
      :::::|||||::: ||| |||||::: ||| ::
21 ThrSerSerSerLeuSerProSerSerSerSerProSerLeuSe 37
      :::::|||||::: ||| |||||::: ||| ::
375 GAGCGGTGCGCGCGACACAGCCGGCTGGTGGAGCCCGGATTGCG 326
      :::: ::::: |||::: ::::: ||| :::
37 rProSerProProSerSerSerProSerAlaProProSerSerLeuS 54
      :||| ::::: |||::: |||::: |||::: |||
325 AACCTGGCCCATCGCTCCCGGACCCACCGAGCGCGCCCGTCACC 276
      :||| ::::: |||::: |||::: |||::: |||
54 erPro.....SerSerProProProLeuSer 62
      :||| ::::: |||::: |||::: |||::: |||
275 GAGCATCCGCAACAGCCCGCGCATCACGAGGGGGCAACCGGCTTTC 226
      :|||:::
63 LeuSerProSer..... 66
      :|||:::
225 GATCAGCTGAGACATCAGCGCGCTGCGGGTCAAGACCCACCTGCGCCAG 176
      :|||::: |||
67 .....SerProProProProp 73
      :|||::: |||
175 GTAGCGACTCCGCGCAGCAGCGCGCGCTGGGGCTGATFCCA 126
      |||::: ||| ::::: |||::: |||::: |||
73 roSerSerSerProLeuSerSerLeuSerSerLeuSerProSerPro 89
      |||::: ||| ::::: |||::: |||::: |||
125 CCAGCCAGCGGAGTGTCGAC.....AGCGGACTGTGCGCGAGCAG 85
      |||::: ||| ::::: |||::: |||::: |||
90 ProSerSerSerProSerAlaProProSerSerLeuSerProSerSe 106
      |||::: ||| ::::: |||::: |||::: |||
84 GCCCATCTGCGCGGCTCTCTGTCGCTGGTGGTGGCGCGCGCGTGGCG 35
      ||| ||| ||| ::::: |||::: |||
106 rPro...ProProLeuSerLeuSerProSerProProProp 122
      |||:::
34 CCACC 30
      ||:::
122 roSer 123
```

seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-7219

```
seq_documentation_block:
; Sequence 7219, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyva, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 7219
; LENGTH: 307
; TYPE: PRT
; ORGANISM: zea mays
US-10-155-881-7219
```

```
Ratio: 1.266      Gaps: 10
Percent Similarity: 50.811  Percent Identity: 27.568

alignment_block:
US-09-462-480-2/rev x US-10-155-881-7219 ..
Align seg 1/1 to: US-10-155-881-7219 from: 1 to: 307

470 GCCCGGTGCGCGGGAAGTCCTGTGTCATTACGGGAGCTCACCAGTCGTC 421
      ||||| ||| ::::: |||::: |||::: |||
19 AlaArgArgAlaProSerSerAlaAlaCysGlyGlyArg.....GI 33
      ||||| ||| ::::: |||::: |||::: |||
420 CTCTTCGTCCTCCAGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 371
      ::::: |||||::: |||::: |||::: |||::: |||
33 yAlaSerSerArgArgArgSerAlaThrArgGlyAlaAlaCysAlaSerG 50
      ::::: |||||::: |||::: |||::: |||::: |||
370 GTCCCGCGCGGACACAGACC.....GGCTGGTGGAG 339
      |||||::: |||
50 lySerAlaProSerThrProArgArgProArgAlaSerThrThrPro 66
      |||||::: |||::: |||
338 CCGCCGATTGCGAACCTCGCCCATCGCTCCCG.....G 304
      |||||::: |||::: |||
67 ProProSer.SerSerAlaGlyProThrProProProThrSerProArgP 83
      |||||::: |||::: |||
303 ACCACCGGAGCGCGCCACCTGTCACCGACGATCCGGCAACAGCCGCCG 254
      ||||| |||::: |||
83 roProProLeuThrAlaAlaProGlyPro.GlyProGlyArgThrArgAr 99
      ||||| |||::: |||
253 GCATCACCAGAGGGGCAACCGGCTTTTCGATCAGCTGAGACATCAGCGGC 204
      ||||| |||::: |||
99 gArgHisArg.....GlnArgAlaThrSerProAlaArgSerArgArgA 114
      ||||| |||::: |||
203 GTCCGGGTCAACGAC.....CCACTGCGCCAGCTAG 172
      ||||| |||::: |||
114 rg.ArgProCysSerAlaArgSerProProCysProProSerProArgTh 130
      ||||| |||::: |||
171 CGACTCCGCGCGCAGCAGCGCGCGCGCTGGGCGCTGATCCACCAG 122
      : ||| |||::: |||
130 rThrProThrArgArgProProSerAlaAlaAlaTrpProSerSerArgA 147
      : ||| |||::: |||
121 CCAGCGGATGGTTCGAC.....AGCGGACTGGTGGCGCAGCAGGCC 81
      ::||| ::::: |||
147 rgArgSerTrpAlaSerSerCysArgSerArgThrArgProSerThrPro 163
      ::||| ::::: |||
80 ATCTGCGCGGCTTCTCTCGTGGCT..... 57
      ::|||::: |||::: |||
164 ProAlaAlaSerGlyThrSerSerProThrArgGlySerSerThrProSe 180
      ::|||::: |||::: |||
56 .....GGTTGCGCGCGCGCGGTGCGCGCCACC 30
      |||::: |||::: |||
180 rArgArgArgArgArgArgArgGlyMetProProProProSerProSer 196
      |||::: |||::: |||
```


OM of: US-09-462-480-2 to: PIR_71.* out_format : pfs
Date: Jul 22, 2002 1:26 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-Q=Cgn2.1/USPTO.spool/US09462480/runat_18072002.164418_19445/app_query.fasta_1.2850
-DB=pir_71 -QSWT=fastan -SUFFIX=rpr -CAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -CAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFWT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09462480_@CGN1_1_73
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMECUT=120 -WARN_TIMECUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-462-480-2
Query length: 524
Database: PIR_71.*
Database sequences: 283138
Database length: 96089334
Search time (sec): 87.620000

score_list:	Strd Orig	zScore	Escore	Len	Documentation
pir2:G70802	+ 713.00	936.10	1.3e-44	368	! probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
pir2:T10033	+ 198.00	261.70	5.8e-07	302	! hypothetical protein - Mycobacterium tuberculosis (strain H37RV)
pir2:S52796	- 140.00	183.48	0.0098	403	! prpL2 protein - human (fragment)
pir2:S20590	- 133.00	168.34	0.0302	913	! exo-alpha-sialidase (EC 3.2.1.1)
pir2:A46302	- 131.00	167.58	0.0430	707	! PTB-associated splicing factor
pir2:T31611	- 131.00	161.70	0.0408	1585	! hypothetical protein Y50E8A.g
pir2:T22602	- 129.50	171.76	0.0595	304	! hypothetical protein F54B11.2
pir1:EM55	+ 127.00	160.90	0.0832	860	! elastin precursor - mouse
pir2:JC4163	+ 126.50	160.55	0.0908	825	! DNA-binding protein 5B5 - rat
pir2:D96711	- 125.50	160.35	0.1095	708	! hypothetical protein F24J5.8 [1]
pir2:S20590	+ 125.50	158.50	0.1067	913	! exo-alpha-sialidase (EC 3.2.1.1)
pir2:BA40505	+ 125.50	152.94	0.1014	1958	! hypothetical protein - suid he
pir2:T35474	- 125.00	162.90	0.1215	456	! 50kD proline rich protein - Str
pir2:A70507	- 124.50	157.44	0.1265	882	! probable-reductase (EC 1.3.99. -
pir2:T35192	- 124.00	158.02	0.1392	744	! probable ABC transporter - Stre
pir2:G70555	- 123.00	165.36	0.1781	227	! hypothetical protein Rv1158c -
pir2:S13367	- 123.00	158.20	0.1669	606	! Om(1D) protein - fruit fly (Dro
pir2:S06733	- 122.50	157.38	0.1813	620	! hydroxyproline-rich glycoprotein
pir2:T13078	- 122.50	155.78	0.1787	772	! KIAA0992 protein - human (fragm
pir2:T50568	- 121.50	150.49	0.2040	1334	! probable multi-domain regulato
pir2:S51939	- 121.00	157.93	0.2386	439	! chitinase (EC 3.2.1.14) precurs
pir2:B39369	- 120.00	155.07	0.2786	543	! homeotic protein barH1 - fruit
pir2:A61183	- 119.50	158.45	0.3143	312	! hypothetical protein (sdsB regi
pir1:BA5344	+ 119.50	145.96	0.2807	1733	! probable nuclear antigen - sui
pir1:WMB338	+ 119.00	156.79	0.3388	358	! infected cell protein ICP34.5 -
pir2:H75457	- 119.00	156.07	0.3366	395	! hypothetical protein - Deinococ
pir2:T33663	- 119.00	155.56	0.3351	424	! hypothetical protein W04B5.3 -
pir2:JQ1696	- 119.00	155.52	0.3350	426	! pistil extensin-like protein pr
pir1:CG9015	- 119.00	148.05	0.3131	1188	! extensin-like protein - maize
pir1:CG8028	+ 118.50	155.29	0.3658	402	! collagen alpha 2(I) chain - bov
pir2:T43556	- 118.50	152.69	0.3573	574	! Wiskott-Aldrich syndrome protei
pir2:T38819	- 118.50	152.69	0.3573	574	! Wiskott-Aldrich syndrome protei
pir2:A56154	- 118.50	151.42	0.3532	684	! Abi substrate ena (enabled) - H
pir2:T75311	+ 118.00	152.32	0.3897	552	! ABC transporter, ATP-binding pr
pir2:T04859	- 118.00	149.27	0.3791	839	! extensin homolog F28A21.80 - Ar
pir2:T46289	- 118.00	149.08	0.3784	862	! hypothetical protein DKF2p434A1
pir2:C86441	- 118.00	146.66	0.3702	1201	! unknown protein [imported] - A
pir2:A41724	- 118.00	146.59	0.3700	1213	! limb deformity (ld) protein - A
pir2:E70766	- 117.50	149.72	0.4165	721	! hypothetical protein Rv2082 - M
pir1:A45344	- 117.00	144.00	0.4327	1446	! immediate-early protein - suid

pir1:WMBEH6 - 117.00 138.29 0.4110 3164 ! UL36 protein - human herpes
pir2:T20807 + 116.50 153.88 0.5179 340 ! hypothetical protein F13A7.1
pir2:B70694 - 116.50 146.79 0.4857 900 ! probable infB - Mycobacteriu
pir2:S54986 - 116.50 146.17 0.4830 980 ! regulatory protein - Emence
pir2:JQ431 - 116.00 153.06 0.5625 348 ! hypothetical 35.5K protein -
seq_name: pir2:G70802
seq_documentation_block:
seqable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70802
R:Colle, S.T.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Blosch, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: G70802
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-368 <COL>
A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL17965.1; PID:e126
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

alignment_scores:	Quality:	713.00	Length:	143
Percent Similarity:	Ratio:	5.057	Gaps:	0
Percent Identity:		98.601	Percent Identity:	96.503
alignment_block:				
US-09-462-480-2 x G70802				
Align seg 1/1 to: G70802 from: 1 to: 368				
1	CTCACAGGTGACGTCGTTGTTTCAGCCAGGTGGGGGGGACCGGCGCGG 50			
226	LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242			
51	CAACCCAGCGCGAGGAGCCGCGAGATGGGCTCTCGGCACCACTC 100			
242	YASNProAlaAspGluAlaGlnMeuGlyLeuGlyThrSerP 259			
101	CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCCCGCGCGCGG 150			
259	roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275			
151	GGCTGTGTCGCGCGGAGTCGCTACCTGGCGGAGTGGGTGCTGTGACCCG 200			
276	GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292			
201	CACGCGCTCATGTCAGCTGATCGAAAGCCGGTGGCCCTCGGTGA 250			
292	gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309			
251	TGCCCGCGGCTGTGGCGGATCGTTCGGTACGGGTGGCGCGCGCGGTG 300			
309	etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325			
301	GGTCCGGGACGATGGCGAGGTTCGCATCCGGGGGCTCCACCAACCC 350			
326	GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342			
351	GGGTCTGGTCCGCGCGACCGCTCGCGGAGGAGCGTGAAGAAAGACGACG 400			
342	oglyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359			
401	AGGACGACTGGGACGAGAGGAGGACTGG 429			

|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: pir2:T10033

seq_documentation_block:
hypothetical protein MLCB628.14c - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C:Accession: T10033
R:Biglmeier, K.; Honore, N.; Woods, S.A.; Caudron, B.; Cole, S.T.
Mol. Microbiol. 7, 197-206, 1993
A:Title: Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae
A:Reference number: Z16917; MUID:93188700
A:Accession: T10033
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <BIG>
A:Cross-references: EMBL:Y14967; NID:g2370268; PIDN:CAA75201.1; PID:g2370281
C:Genetics:
A:Note: MLCB628.14c

alignment_scores:
Quality: 198.00 Length: 126
Ratio: 2.329 Gaps: 4
Percent Similarity: 67.460 Percent Identity: 41.270

alignment_block:

US-09-462-480-2 x T10033 ..

Align seg 1/1 to: T10033 from: 1 to: 302

1 CTGCAGCAGGTGACGTCTGTTTCAGCAGGTGGCGGCACCGCGGG 50
190 MetGlnGlnVallySerLeuPheThrSerIleAspSerThrGlyValty 206
51 CAACCCAGCGCAC.....GAGGAAGCGCGCAGATGGCGCTGC 88
206 rThrSerAlaGlnArgGlyAspThrGluSerAlaHisArgIleGlyLeup 223
89 TCAGCAGCAGTCCGCTGTCACACCATCCGTGGCTGGTGTGATCAGGCC 138
223 heGlyAlaSerThrLeuSerSerHisProLeuValGlyIleThrGlyThr 239
139 AGCGCGGCGCGCGCTGCTGCGCGGAGTCGCTACCTGCGCAGGTGG 188
240 ThrThrAspThrArgLeuLeuCysAlaGluSerLeuProSerAlaSerGI 256
189 GTGCTTGACCCGCGCGCTGATGCT.....CAGCTGATCGAAAGC 232
256 ySerLeuAlaTrpThrProLeuMetThrGlnPheGlnLeuIleAspLys 273
233 CGGTGCGCCCTCG.....GTGATGCGCGCGGTGTGTC 267
273 erIleAlaProGluProArgGlnArgValMetLeuProTrpAlaAla 289
268 GGATGCGCGGTGACGGGTGCGCGCTGCGGTGGTCCGGGAGCGATGG 317
290 GlySerPro.....G 293
318 CCAGGGTTCGCAATCGCGCGGTCCAC 345
293 yHisAsnAlaGlnAspGlyGlyThrThr 302

seq_name: pir2:S52796

seq_documentation_block:
pripL2 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 19-May-1995 #sequence_revision 10-Nov-1995 #text_change 10-Sep-1997
C:Accession: S52796
R:Ruhmann, A.; Kreidewiss, S.; Nordheim, A.

submitted to the EMBL Data Library, March 1995
A:Reference number: S52796
A:Accession: S52796
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-403 <RUH>
A:Cross-references: EMBL:X86019; NID:g762950; PID:g762951

alignment_scores:
Quality: 140.00 Length: 163
Ratio: 2.000 Gaps: 9
Percent Similarity: 42.945 Percent Identity: 32.515

alignment_block:

US-09-462-480-2/rev x S52796 ..

Align seg 1/1 to: S52796 from: 1 to: 403

472 CGGCCCGGGTGGCGGGAAGTCTCTTGTTCATTACGGGAGCTCACCAGTCG 423
|||||
224 ArgProGlyProPro..... 228
422 TCCTCTTCGTCCTCCAGTCGT.....CCTC 400
|||||
229 .ProLeuProProSerSerSerGlyAsnAspGluThrProArgLeuProG 245
399 GTCGTCTTCTACGCTCCCTGCGGAGCGGTGCGCGGCGGACGACGACCG 350
|||||
245 lnArgAsnLeu....SerLeuSerSerSerThrProLeuProSerPro 260
349 GGCTGGTGGAGCCCGCGGATTGCGAACCTGCGCCATCGCTCCCGGACCC 300
|||||
261 Gly.....ArgSerGlyPro...LeuProProPr 269
299 ACCGAGCGGGCGCACCGCTACCGACGATCCGCGACACGACGCGCGGAT 250
|||||
269 oValProSerGluArgProProProValArgAspProGlyArgS 286
249 CA.....CCGAGGGGCAACCGCTTTTCGATCAGCTGAGCATCA 209
|||||
286 erGlyProLeuProProProProValSerArgAsnGlySerThrSer 302
208 CGCGCGTGGGGTCAACGACCCACCTCGCCA.....GGTAGCGAC 168
|||||
303 ArgAlaLeuProAlaThr..ProGlnLeuProSerArgSerGlyValAsp 318
167 TCCGCGCGCAGCAGCGCGCGCGCTGGGCGCTGATCCACCAGCCAG 118
|||||
319 SerProArgSerGlyProArgProProLeuProProAspArgPro.... 333
117 CGGATGGTTCGACAGCGGACTGGTGCCGAGCAGGCCCATCTGCGCGGCTT 68
333

seq_name: pir2:S20590

seq_documentation_block:
exo-alpha-sialidase (EC 3.2.1.18) - Actinomyces viscosus
C:Species: Actinomyces viscosus
C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 22-Oct-1999
C:Accession: S20590
R:Henningsen, M.; Roggentin, P.; Schauer, R.
Biol. Chem. Hoppe-Seyler 372, 1065-1072, 1991

A:Title: Cloning, sequencing and expression of the sialidase gene from Actinomyces viscosus
A:Reference number: S20590; MUID:92162190
A:Accession: S20590
A:Status: preliminary
A:Molecule type: DNA

A;Residues: 1-913 <HEI>
A;Cross-references: EMBL:X62276; NID:g39254; PIDN:CAA44166.1; PID:g39255
C;Keywords: glycosidase; hydrolase

alignment_scores:
Quality: 133.00 Length: 188
Ratio: 1.529 Gaps: 9
Percent Similarity: 46.277 Percent Identity: 29.787

alignment_block:
US-09-462-480-2/rev x S20590 ..

Align seg 1/1 to: S20590 from: 1 to: 913

```
472 CGGCCGGGTGGCGGGAAGTCTGTGTCATTACGGAGTCCACCAGTCG 423
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
686 ArgProArgArgProArgArgAlaLeuSerProArgArgHisArgHisI 702
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 TCCTCTTCGTCCTCCAGTCCTCGTCGTCCTTCTTCCAGCTCTCGCGGAG 373
:|||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
702 sProArgProSerArgAlaLeuArg.....ProSerArgA 715
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 CGGTGCGCGG.....CGACCAGACCGGGGTGTGTGGAGC 338
|| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
715 laGlyProGlyAlaGlyAlaHisAspArgSerGluHisGlyAlaHisThr 731
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 CGCGGATTGGAACCTCGGCCCATCGTCCCGGACCCACCGAGCGCG 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
732 GlySerCysAlaGlnSerAlaProGluGlnThrAspGlyProThrAlaAl 748
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 CCACCGGTACCGAGCATCCCGGCAACAGCCCGGCATCACCGAGGGGCG 238
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
748 a...ProAlaProGluThrSerSerAlaProAlaAlaGluProThrGlnA 764
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 AACCGGCTTTTCGA.....TCAGCTGAG 215
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
764 laProThrValAlaProSerValGluProThrGlnAlaProGlyAlaGln 780
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 ACATCAGCGCG.....TGCGGTCAACGACCCACCTGCGCCAGGTAGC 171
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
781 ProSerSerAlaProLysProGlyAlaThrGlyArgAlaProSerValva 797
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
170 GACTCCGCGCGCAGCAGC..... 152
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
797 laSnProLysAlaThrGlyAlaAlaThrGluProGlyThrProSerSers 814
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 .....CGCGCGCGCGTGGGCG 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
814 erAlaSerProAlaProSerArgAsnAlaAlaProThrProLysProGly 830
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
133 CTG.....ATCCACGACGCGGATGTTCCGACAGCGGACTG 96
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
831 MetGluProAspGluIleAspArgProSerAspGlyThrMetAlaGlnPr 847
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 GTGCGGACGAGCCCATCTGCGCGCTTCCTCGTGGCTGGGTGCCGCC 46
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
847 oThrGlyAla...ProAlaArgValProArg.....ArgA 860
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45 GCCGGTGGCGGCCA 32
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
860 rGArgArgArgPro 864
```

seq_name: pir2:A46302

seq_documentation_block:

FTB-associated splicing factor, long form - human
N;Alternate names: myoblast cell surface antigen 24.1D5; polypyrimidine tract-binding protein
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C;Accession: A46302; A43557; S29995
R;Patton, J.G.; Porro, E.B.; Galceran, J.; Tempst, P.; Nadal-Ginard, B.
Genes Dev. 7, 393-406, 1993

A;Title: Cloning and characterization of PSF, a novel pre-mRNA splicing factor.
A;Reference number: A46302; MUID:93194059
A;Accession: A46302
A;Molecule type: mRNA
A;Residues: 1-707 <PAT>
A;Cross-references: EMBL:X70944; NID:g38457; PIDN:CAA50283.1; PID:g38458
A;Note: sequence extracted from NCBI backbone (NCBIP:127206)
R;Gower, H.J.; Moore, S.E.; Dickson, G.; Elsom, V.L.; Nayak, R.; Walsh, F.S.
Development 105, 723-731, 1989
A;Title: Cloning and characterization of a myoblast cell surface antigen defined by
A;Reference number: A43557; MUID:90091812
A;Accession: A43557
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 312-707 <GOW>
A;Cross-references: GB:X16850; NID:g23711; PIDN:CAA34747.1; PID:g23712
C;Genetics:
A;Gene: GDB:SFPQ; PSF
A;Cross-references: GDB:138275
A;Map position: 4q-4q
C;Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleop
C;Keywords: alternative splicing; pre-mRNA splicing; surface antigen
F;298-359/Domain: ribonucleoprotein repeat homology <RRM1>
F;372-438/Domain: ribonucleoprotein repeat homology <RRM2>

alignment_scores:
Quality: 131.00 Length: 139
Ratio: 1.899 Gaps: 9
Percent Similarity: 49.640 Percent Identity: 31.655

alignment_block:

US-09-462-480-2/rev x A46302 ..

Align seg 1/1 to: A46302 from: 1 to: 707

```
367 CCGCGCGCAGCAGACCCGGGTGTGGAGCCCGCGGATTGCGAACCTGG 318
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
46 ProMetGlyProGlyProGlyGln.....SerGlyProly 57
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 CCCATCGTCCGGACCCACCG.....GAGCGCGCCACCGTCAC 277
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 sProProileProProProProProHisGlnGlnGlnGlnProProp 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 CGACGATCCGCGCACAGCCGCGGCATCA.....CCGAGGGGCGCAACCG 233
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
74 ro.....GlnGlnProProProGlnGlnProProHisGlnPro 87
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 GCTTTTCATCAGTCGACATCAGCGGTGCGGGTCAACGACCCACT 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
88 ProProHisProGlnPro.HisGlnGln.....GlnGln.ProPro 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
182 GCGCCAGGTAGCTCCGCGCGC.....AGCAGGCC 151
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 ProProProGlnAspSerSerLysProValAlaGlnGlyProGlyPr 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
150 CGCGCCCGCGTGGGGCTGATCCACCGACCGCGGA..... 114
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 oAlaProGlyValGlySerAlaProProAlaSerSerSerAlaProProA 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
113 .....TGGTTCGACGCGGACTGTG 93
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 laThrProProThrSerGlyAlaProProGlySerGlyProGlyProThr 150
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
92 CCGAGCAGGCCCATCTCGCGGCTTCCTCGTGGGTGGTCCGCCGCC 43
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ProThrProProAlaValThrSerAlaProProGlyAlaProProPr 167
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
42 GGTGCCGCCACC 30
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 oThrProProSer 171
```

seq_name: pir2:T31611

seq_name: pir1:EAMS

seq_documentation_block:

elastin precursor - mouse

N:Alternate names: tropoelastin

C:Species: Mus musculus (house mouse)

C:Date: 18-Aug-1995 #sequence_revision 16-Aug-1996 #text_change 22-Jun-1999

C:Accession: A55721

R:Wydner, K.S.; Sechler, J.L.; Boyd, C.D.; Passmore, H.C.

Genomics 23, 125-131, 1994

A:Title: Use of an intron length polymorphism to localize the tropoelastin gene to mouse

A:Reference number: A55721; MUID:95130069

A:Accession: A55721

A:Molecule type: mRNA

A:Residues: 1-860 <WYD>

A:Cross-references: GB:U08210; NID:9473273; PIDN:AMA80155.1; PID:9473274

C:Genetics:

A:Map position: 5

C:Superfamily: elastin

C:Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxyllysine

F:1-27/Domain: signal sequence #status predicted <SIG>

F:28-860/Product: elastin #status predicted <WAT>

F:850-855/Disulfide bonds: #status predicted

alignment_scores:

Quality: 127.00 Length: 141

Ratio: 1.649 Gaps: 8

Percent Similarity: 54.610 Percent Identity: 34.752

alignment_block:

US-09-462-480-2 x EAMS ..

Align seg 1/1 to: EAMS from: 1 to: 860

31 GTGGGGCCACCGG.....GCGGCAACCC 56
:::||||| |||
53 IleglyGlyLeuGlyGlyGlyAlaLeuGlyProGlyGlyLysPr 69
57 AGCGACGAGGAGCGCGCAGATGGCGCTCTCGGCACCACTCGCTGT 106
69 oProLysProGlyAlaGlyLeuLeuGlyThrPheGlyAlaGlyProGlyG 86
107 CGAACCATCCGCTGGCTGGATCAGCCCGCGCGCGCGCGCGCTG 156
::: |||:::|||||:::|||||::: |||
86 ly.....LeuGlyGlyAlaGlyProGlyAlaGlyLeuGlyAla 98
157 CTGCGCGGAGTGCCTACCTGCGCGCAGGTGGTGGTGCACCG..... 200
::: |||:::|||||:::|||||::: |||
99 PheProAlaGlyThrPheProGlyAlaGlyAlaLeuVal..ProGlyGly 114
201CACGCGCTGATGCTCAGCTGATCGAAAGCGGTGCCCC 241
||||| |||::: ||| |||::: |||
115 AlaAlaGlyAlaAlaAlaTyrLysAlaAlaLysAlaGlyAlaGl 131
242 CCTCGGTGATCCGCGCTGCTGCCGA.....TCGTCGGTG 279
|||||::: |||::: ||| |||
131 yLeuGlyGlyValGlyGlyValProGlyGlyValGlyValGlyVal 147
280 ACGGTGCGCGCGCTCGGTGGGT...CCGGAGCGATGGGCCAGGGTTC 326
|||||::: ||| |||::: ||| |||
148 ProGlyGlyValGlyValGlyGlyValProGlyGlyValGlyValGly 164
327 GCATCCGCGCGC.....TCCACCA 346
::: |||::: |||::: |||
164 yValProGlyGlyValGlyGlyIleGlyGlyIleGlyGlyValGlyVal 181
347 GCCCGGCTCGTCCGCGCG 366
|| ||| |||::: |||
181 erThrGlyAlaValPro 187

seq_name: pir2:JC4163

seq_documentation_block:

DNA-binding protein SE5 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 27-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 20-Jun-2000

C:Accession: JC4163; PC4040

R:Suzuki, E.; Kojima, N.; Yoshimura, K.; Obata, K.; Akagawa, K.

J. Biochem. 118, 122-128, 1995

A:Title: Cloning and sequence analysis of cDNA for a possible DNA-binding protein SE5

A:Reference number: JC4163; MUID:96015159

A:Accession: JC4163

A:Molecule type: mRNA

A:Residues: 1-825 <SUZ>

A:Cross-references: DDBJ:D37934; NID:g531260; PIDN:BAA07153.1; PID:g531261

A:Experimental source: brain

A:Accession: PC4040

A:Molecule type: protein

A:Residues: 230-455 <SUZ>

C:Comment: This protein has an abundance of arginine, a glycine-rich region and a pro

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homol

C:Keywords: nerve; phosphoprotein

F:436-443/Region: nuclear location signal

F:722-731/Region: proline cluster

F:62,258,345,360,404,413,570,613,635,752,820/Binding site: phosphate (Ser) (covalent

alignment_scores:

Quality: 126.50 Length: 183

Ratio: 1.265 Gaps: 8

Percent Similarity: 54.645 Percent Identity: 32.240

alignment_block:

US-09-462-480-2 x JC4163 ..

Align seg 1/1 to: JC4163 from: 1 to: 825

24 CAGCCAGGTGGG.....CGGCACCGCGCGGCAACCCACGCCGACGAGG 67
478 GlnAlaGlyGlyGlyProArgGlyArgAlaArgGlyProArgGlnGlnAl 494
68 AAGCGCGCAGATGGCGCTGCTCGGCACCACTCGCTCTCGAACCATCCG 117
494 aArgArgHisGlyProGlnArgArgGly.....Pro.Pro 507
118 CTGCTGTGTGATCAGCGCCCGCGCGCGCGCTGCTG..... 159
|||||::: |||::: |||::: |||
508 GlnAlaGlyGlyGlyProGlyAspAlaThrLeuValLeuGlyLeuGl 524
160CGCGCGAGTCGTACTGCGCG 181
524 yThrThrSerGlyGluGlnArgAlaAspGlnSerGlnThrLeuProAlaL 541
182 CAGTGGGTGCTGACCCG.CACGCCGCTGATGCTCAGCTGATCGAATAA 230
::: |||::: ||| |||::: |||
541 euAlaGlyAlaProThrAlaHisAlaHisAlaValProGlyProGlyPro 557
231 GCCGGTTGCCCGCTCGGTGATCGCGC.....GGCTGTCGCCGATCGT 274
|||||::: |||::: ||| |||
558 AlaAlaThrLeuGlyGlyArgGlyArgGlySerTrpArg..... 572
275 CGGTGACGGGTGGCGCGCTCGGTGGGTCCGGAGGAGGATGGGCCAGG 324
||| |||::: |||::: |||::: |||
573GlyGlyArgGlyGlyGlyAlaGlyAlaSerGlyGlyLeu 587
325 TCGCAATCCCG.....CGGCTCCACACGCCCGGCTGTGTCGCCGCC 365
::: |||::: |||::: |||
587 rgGlyGlyArgGlyArgGlyGlyArgGlyArgGlySerGlyLeuSer 603
366 GGCACCGCTCGC...GCAGGAGCGTGAAGAGACGACGACGACGCTGG 412
|||||::: |||::: |||::: |||
604 GlyThrArgGluAspAlaGlySerProSerAlaArgGlyGluGlnAr 620
413 ACGAAGAGGACGACTGGTGGTGGTCCCGTAATGACAACAGACTTCGCCGCC 462


```

437 GGAGCTCACCAGTGTCTCTCTGTCCTCCAGTCGTCTCTCTGTCGTCTCTCTTC 388
||||| : : : : : ||| : : : : : : : : : :
293 GlyAlaGlyGluIleProAlaAlaArgGlnGlnGluArgArgGluAlaAs 309
: : : : : ||| : : : : : ||| : : : : :
387 ACGETCTGCGGAGCGGTGCGGC.....GCGACGAGACCGGGC 347
: : : : : ||| : : : : : ||| : : : : :
309 pGlyGluArgAlaProGlyThrGlyThrLeuSerAlaGlyArgProGly. 325
||| : : : : : ||| : : : : : ||| : : : : :
346 TGTGTGAGCGCGCGGATTGCGAACCTGCGCCCATCGCTCCCGGACCCACC 297
||| : : : : : ||| : : : : : ||| : : : : :
326 .....ProGlu...ThrProGlyValLeuLeuProSerProAla 337
296 GGAGCGCGCCACCGCTCACCGACGATCCGGCAACAGCCCGCGCATCAC 247
: : : : : ||| : : : : : ||| : : : : :
338 ThrGluGlyGluProLeuGluAspGlyArgGlyAlaGlyHisGlyAspG1 354
||| : : : : : ||| : : : : : ||| : : : : :
246 CGAGGGGGCAACCGGCTTTTCGATCAGCTGAGACATCAGCGGGGTGCGGG 197
: : : : : : : : : : : ||| : : : : :
354 yAspGlyProAlaSerValAlaValProAsnArgThrProGlyProArg. 370
||| : : : : : ||| : : : : : ||| : : : : :
196 TCAACGACCCACCTGCGCCA.....GGTAGCGACTCGCGCGC 159
: : : : : ||| : : : : : ||| : : : : :
371 .....GlnAlaProAlaProValSerGlyHisGlyProGluAlaAlaPro 385
158 AGCAGGCGCGCGCGCTGGGGCTGTATCCACCGAGCGGATGGTT 109
||| : : : : : ||| : : : : : ||| : : : : :
386 SerProSerAlaProAlaProGlyProSerGluProAlaSerGly.... 400
108 CGACAGCGGACTGGTCCGAGCAGGCCCATCTGCGGGCTTCTCTCGTCGG 59
||| : : : : : ||| : : : : : ||| : : : : :
401 .ProSerAlaProAlaProGlyProProAlaProAlaAlaGlyProSerA 417
58 CTGGGTTGCGCGCGCGGTGCGGCC 33
||| : : : : : ||| : : : : :
417 laProAlaProGlyProSerAlaPro 425

```

THIS PAGE BLANK (USPTO)

score_list:	Sequence	Strd	Orig	zScore	EScore	Len	Documentation
	SwissProt_40:WAIP_HUMAN		145.50	158.64	0.0768	503	! Q43516 homo sapiens (human)
	SwissProt_40:GPI_CHLRE		131.50	143.47	0.4872	555	! Q9f96g chlamydomonas reinhardtii
	SwissProt_40:SFQO_HUMAN		131.00	141.31	0.5041	707	! P23246 homo sapiens (human)
	SwissProt_40:ELS_MOUSE		127.00	135.84	0.8356	860	! P54320 mus musculus (mouse)
	SwissProt_40:SE5_RAT		126.50	135.61	0.8980	825	! Q63003 rattus norvegicus (rat)
	SwissProt_40:KLP2_HUMAN		125.00	139.75	1.23	355	! Q9Y5W3 homo sapiens (human)
	SwissProt_40:HM1D_DROAN		123.00	134.07	1.49	606	! P23544 drosophila ananassae
	SwissProt_40:EXTN_TOBAC		122.50	133.39	1.59	620	! P13983 nicotiana tabacum (coca
	SwissProt_40:TPM5_DROME		121.50	133.76	1.86	504	! P49456 drosophila melanogaster
	SwissProt_40:T2D3_HUMAN		120.50	127.55	1.92	1083	! Q00268 homo sapiens (human)
	SwissProt_40:DIAL_HUMAN		120.50	126.59	1.89	1248	! Q60610 homo sapiens (human)
	SwissProt_40:VNUA_PVKVA		119.50	123.34	2.06	1733	! P33485 pseudorabies virus (c
	SwissProt_40:EXLP_TOBAC		119.00	132.30	2.66	426	! Q03211 nicotiana tabacum (coca
	SwissProt_40:FMN_CHICK		118.00	124.20	2.64	1213	! Q05858 gallus gallus (chicken)
	SwissProt_40:HK61_RAT		117.50	131.79	3.31	365	! Q35672 rattus norvegicus (rat)
	SwissProt_40:FK82_MYCTO		117.50	127.19	3.02	721	! Q10690 mycobacterium tuberculosis
	SwissProt_40:IE18_PVKVA		117.00	121.97	2.94	1446	! P33479 pseudorabies virus (c
	SwissProt_40:TEGU_HSV11		117.00	116.68	2.65	3164	! P10220 herpes simplex virus (h
	SwissProt_40:1F2_MYCTO		116.50	124.66	3.35	900	! P71613 mycobacterium tuberculosis
	SwissProt_40:YV35_STRFR		116.00	130.56	4.07	348	! P20186 streptomyces fradiae
	SwissProt_40:SEPA_EWENI		116.00	119.50	3.27	1790	! P78621 emericella nidulans
	SwissProt_40:AFSK_STRGR		115.50	124.36	3.88	807	! P54742 streptomyces griseus
	SwissProt_40:TCOF_HUMAN		115.50	120.58	3.61	1411	! P54742 streptomyces griseus
	SwissProt_40:APG_ARATH		115.00	126.63	4.39	534	! Q13428 homo sapiens (human)
	SwissProt_40:IE18_PVKF		115.00	119.83	3.84	1461	! Q40602 arabidopsis thaliana
	SwissProt_40:MYCN_MARMO		114.00	126.60	5.11	460	! P11675 pseudorabies virus (c
	SwissProt_40:S3A2_HUMAN		114.00	126.55	5.10	464	! Q15428 homo sapiens (human)
	SwissProt_40:EBN1_EBV		113.50	123.84	5.23	641	! P03211 Epstein-Barr virus (str
	SwissProt_40:VRP1_YEAST		113.50	122.20	5.06	817	! P37370 saccharomyces cerevisiae
	SwissProt_40:CA21_RANCA		113.50	118.79	4.73	1355	! Q42350 rana catesbeiana (bull
	SwissProt_40:WASP_MOUSE		113.00	124.74	5.74	520	! P70315 mus musculus (mouse)
	SwissProt_40:CAPU_DROME		112.50	119.42	5.58	1059	! Q24120 drosophila melanogaster
	SwissProt_40:DRPL_RAT		112.50	118.67	5.50	1183	! P54528 rattus norvegicus (rat)
	SwissProt_40:FM14_MOUSE		112.50	118.54	5.49	1206	! Q05859 mus musculus (mouse)
	SwissProt_40:CA21_BOVIN		112.50	117.71	5.40	1364	! P02465 bos taurus (bovine)
	SwissProt_40:FMN1_MOUSE		112.50	117.21	5.34	1468	! Q05860 mus musculus (mouse)
	SwissProt_40:NK1_RAT		112.50	117.09	5.33	1493	! Q62925 rattus norvegicus (rat)
	SwissProt_40:D4DR_HUMAN		112.00	124.43	6.65	467	! P21917 homo sapiens (human)
	SwissProt_40:SYN1_MOUSE		112.00	121.99	6.34	670	! Q88935 mus musculus (mouse)
	SwissProt_40:SYN1_HUMAN		112.00	121.65	6.30	705	! Q10960 homo sapiens (human)

seq_name: SwissProt 40:WAIP_HUMAN

```
FT VARIANT 495 495 G -> A.
FT CONFLICT 303 309 PHRPLR -> SSOAPPP (IN REF. 3).
FT CONFLICT 360 360 P -> PV (IN REF. 2).
FT CONFLICT 487 503 GSNRRRGPPPLPIPR -> EYFCOGF (IN REF. 2).
SQ SEQUENCE 503 AA; 51489 MW; 43EB88674DD3BF1A CRC64;

alignment_scores:
  Quality: 145.50 Length: 163
  Ratio: 2.021 Gaps: 10
  Percent Similarity: 44.172 Percent Identity: 33.742

alignment_block:
US-09-462-480-2/rev x WAIPI_HUMAN ..
Align seg 1/1 to: WAIPI_HUMAN from: 1 to: 503

472 CGGCGCGGTGGCGGGGAAGTCTGTGTCATTCGGGAGTCACAGTCG 423
||||| |||
315 ArgProGlyProPro..... 319
422 TCCTCTTCCTCCAGTCGT.....CCTC 400
320 .ProleuProProSerSerSerGlyAsnAspGluThrProArgLeuProG 336
399 GTCGTCTTCTTCACGCTCCTCGCGAGCGGTGCGCGCGGCGACACGCG 350
336 InArgAsnLeu...SerLeuSerSerThrProProLeuProSerPro 351
349 GCCTGGTGGAGCCCGCGGATTCGCAACCTGGCCCATCGCTCCGGACCC 300
||| : : : : : ||| |||
352 Gly.....ArgSerGlyPro...LeuProProPr 360
299 ACCGAGGCGGCGCACCGCTCACCGAGTCCGGCAACAGCGCGCGCAT 250
||||| ||| ||| : : : : : |||
360 OProSerGluArg...ProProProValArgAspProGlyArgS 376
249 CA.....CCGAGGGGCAACCGCTTTCATCAGCTCAGACATCA 209
||| ||| ||| ||| ||| ||| |||
376 erGlyProProProProProProValSerArgAsnGlySerThrSer 392
208 GCGGCGTGGGTCAACGACCCACCTCGGCA.....GGTAGGAC 168
||| : : : : : ||| ||| |||
393 ArgAlaLeuProAlaThr...ProGlnLeuProSerArgSerGlyValAsp 408
167 TCCGCGCGAGCAGCGCGCGCGCTGGCTGATCCACCGACGAC 118
||| ||| ||| ||| ||| ||| |||
409 SerProArgSerGlyProArgProProLeuProProAspArgPro..... 423
117 CGGATGTTTCGACAGCGGACTGTTGCCGAGCAGGCCCATCTGCGCGGCTT 68
423 ..... 423

seq_name: SwissProt_40:GP1_CHLRE

seq_documentation_block:
ID GP1_CHLRE STANDARD; PRT; 555 AA.
AC Q9FP06; Q03927;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gpi precursor (hydroxyproline-rich
DE glycoprotein 1).
GN GP1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
```



```

|||||:  |||||||:  |||||||:
107 aProSerProAlaProSerProAla.....ProProS 120
252 CATCACGAGGGGCAACCGGCTTTTCATCAGCTGAGATGAGATGAGCGCG 203
    :  |||  |||  |||  |||
120 erProProSerProAlaProPro..SerProSer..... 130
202 TGGCGGTCAACGACCCACCTCGCCAGGTAGCGACTCCGCGCGCAGCAGG 153
    |||||||  ||  |||  |||
131 .....ProProAlaProProSerProSerProSer... 141
152 CCGCGCGCGCGCTGGGCGCTGATCCA....CAGCCAGCGGATGGTTCGA 106
    |||||||  ||  |||:||||  |||  |||
142 ProAlaProProLeuProProSerProAlaProProSer..... 154
105 CAGCGGACTGGTCGAGCAGCGCCATCTCGCGGGCTTCCTCGTCGCGCTG 56
    ||||||  |||:||||  |||
155 .....ProSerProProValProProSerProSer..... 164
55 GGTGCGCGCGCGGTGGCGCCGCCACC 30
165 .....ProProValProProSer 170

```

seq_name: SwissProt_40:SFPO_HUMAN

seq_documentation_block:

```

ID SFPO_HUMAN STANDARD; PRT; 707 AA.
AC P23246; P30808;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1995 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor, proline and glutamine-rich (Polypyrimidine tract-
DE binding protein-associated splicing factor) (PTB-associated splicing
DE factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).
GN SFPO OR PSF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal brain;
RX MEDLINE=93194059; PubMed=8449401;
RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
RT factor."
RL Genes Dev. 7:393-406(1993).
RN [2]
RP SEQUENCE OF 312-707 FROM N.A.
RC TISSUE=Fetal skeletal muscle;
RX MEDLINE=90091812; PubMed=2480877;
RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
RT "Cloning and characterization of a myoblast cell surface antigen
RT defined by 24 LD5 monoclonal antibody."
RL Development 105:723-731(1989).
RN [3]
RP SEQUENCE OF 48-68 AND 213-246.
RX MEDLINE=93176127; PubMed=8439294;
RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
RT "Purification and characterization of a DNA-binding heterodimer of 52
RT and 100 kDa from HeLa cells."
RL Biochem. J. 290:267-272(1993).
CC -1- FUNCTION: ESSENTIAL PRE-MRNA SPLICING FACTOR REQUIRED EARLY IN
CC SPICEOSOME FORMATION. BINDS TO THE POLYPYRIMIDINE POLYPYRIMIDINE
CC TRACTS. FORMS A COMPLEX WITH THE POLYPYRIMIDINE TRACT-BINDING
CC PROTEIN (PTB). SEEMS TO ALSO BIND DNA.
CC -1- SUBUNIT: HETEROTETRAMER OF TWO 52 KDA AND TWO 100 KDA SUBUNITS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL
CC SURFACE ANTIGEN 24.LD5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN

```

ECTOKINASE.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

```

DR EMBL; X70944; CAA50283.1; -
DR EMBL; X16850; CAA34747.1; -
DR PIR; A43557; A43557.
DR PIR; S29770; S29770.
DR HSSP; P19339; 1SXL.
DR MIM; 605199; -
DR InterPro; IPR000504; RRM.
DR Pfam; PF00076; Rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
KW Alternative splicing
FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
FT REPEAT 9 27 3 X 3 AA REPEATS OF R-G-G.
FT REPEAT 19 11 1.
FT REPEAT 25 21 2.
FT REPEAT 25 27 3.
FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
FT DOMAIN 10 15 POLY-GLY.
FT DOMAIN 20 27 POLY-GLY.
FT DOMAIN 56 65 POLY-PRO.
FT DOMAIN 67 71 POLY-GLN.
FT DOMAIN 95 98 POLY-GLN.
FT DOMAIN 99 103 POLY-PRO.
FT DOMAIN 184 188 POLY-PRO.
FT DOMAIN 571 574 POLY-ARG.
FT DOMAIN 613 616 POLY-GLY.
FT DOMAIN 635 641 POLY-GLY.
FT VARSPIC 663 707
FT CONFLICT 243 243
FT SEQUENCE 707 AA; 76149 MW; 6D8D5EA95E235847 CRC64;
SQ

```

alignment_scores:

Quality:	131.00	Length:	139
Ratio:	1.899	Gaps:	9
Percent Similarity:	49.640	Percent Identity:	31.655

alignment_block:

US-09-462-480-2/rev x SFPO_HUMAN ..

Align seg 1/1 to: SFPO_HUMAN from: 1 to: 707

```

367 CCGCGCGGACGACGACCGCGGCTGGTGGAGCGCGCGGATTCGAAACCTGG 318
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
46 ProMetGlyProGlyProGlyGln.....SerGlyProLy 57
317 CCCATCGCTCCCGGACCCACCG.....GAGCGGCGCCACCGCTAC 277
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
57 sProProileProProProHisGlnGlnGlnGlnGlnProProp 74
276 CGAGATCCGCGCACACGCGCGCGCATCA.....CCGAGGGGCAACCG 233
    ||  |||  |||  |||  |||  |||  |||  |||  |||  |||
74 ro.....GlnGlnProProGlnGlnGlnGlnGlnGlnPro 87
232 GCTTTTCATCAGTCAGATCAGCGCGGTGCGCGGTCAACGACCCACCT 183
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
88 ProProHisProGlnProHisGlnGln.....GlnGlnProPro 100
182 GCGCCAGTAGCAGACTCCGCGCGC.....AGCAGGCC 151

```

```

101 ProProProGlnAspSerSerLysProValValAlaGlnGlyProGlyPr 117
150 CGCGCCGCGCTGGCGGTGATCCACAGCAGCAGCGGA..... 114
117 oAlaProGlyValGlySerAlaProProAlaSerSerAlaProProA 134
113 .....TGGTTCGACAGCGGACTGGTG 93
134 laThrProProThrSerGlyAlaProProGlySerGlyProGlyProThr 150
92 CCGAGCAGGCCATCTCGCGGCTTCCTCGCTGGGTGGCGCGGCC 43
151 ProThrProProProAlaValThrSerAlaProProGlyAlaProProPr 167
42 GGTGCCGCCACC 30
167 oThrProProSer 171

```

seq_name: SwissProt_40:ELS_MOUSE

```

seq_documentation_block:
ID ELS_MOUSE STANDARD; PRT; 860 AA.
AC P54320.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Elastin precursor (Tropoelastin).
GN ELN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Lung;
RX MEDLINE=95130069; PubMed=7829060;
RA Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT "Use of an intron polymorphism to localize the tropoelastin gene to
RT mouse chromosome 5 in a region of linkage conservation with human
RT chromosome 7.";
RL Genomics 23:125-131(1994).
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC NUCAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC INTO AN EXTENSIBLE 3D NETWORK.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U08210; AA80155.1; .
CC DR HSSP; P04002; LWFA.
CC DR MGD; MGI:95317; Eln.
CC DR InterPro; IPR003979; tropoelastin.
CC DR PRINTS; PR01500; TROPOELASTIN.
CC DR Structural protein; Repeat; Signal; Connective tissue.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 860 ELASTIN.
SQ SEQUENCE 860 AA; 71955 MW; 0C0BESAEE1EDD7F1 CRC64;

```

alignment_scores:
 Quality: 127.00 Length: 141
 Ratio: 1.649 Gaps: 8
 Percent Similarity: 54.610 Percent Identity: 34.752

alignment_block:
 US-09-462-480-2 x ELS_MOUSE

Align seg 1/1 to: ELS_MOUSE from: 1 to: 860

```

31 GTGGCGCGCACCGCG.....GGCGGCAACCC 56
   ::::: ||| ||| ||| ||| |||
53 lIeGlyGlyLeuGlyGlyGlyGlyAlaLeuGlyProGlyGlyLysPr 69
   | | | | | | | | | | | | | | | | | |
57 AGCCGACGAGGAAGCGCGCATGCTCGGCACCGACGTCGCGTGT 106
   | | | | | | | | | | | | | | | | | |
69 oProLysProGlyAlaGlyLeuGlyThrPheGlyAlaGlyProGlyG 86
   | | | | | | | | | | | | | | | | | |
107 CGAACCATCCGCTGGCTGGTGGATCAGCCCGCGCGCGCGCGCGCTG 156
   :: | | | | | | | | | | | | | | | | |
86 lY.....LeuGlyGlyAlaGlyProGlyAlaGlyLeuGlyAla 98
   | | | | | | | | | | | | | | | | | |
157 CTGCGCGCGGAGTCGTACCTGGCGCAGGTGGTGGTGGTACCCG..... 200
   :: | | | | | | | | | | | | | | | | |
99 PheProAlaGlyThrPheProGlyAlaGlyAlaLeuVal..ProGlyGly 114
   | | | | | | | | | | | | | | | | | |
201 .....CACGCGCTGATCTCTCAGCTGATCGAAAGCGGTGGCC 241
   | | | | | | | | | | | | | | | | | |
115 AlaAlaGlyAlaAlaAlaAlaTyLysAlaAlaLysAlaGlyAlaGl 131
   | | | | | | | | | | | | | | | | | |
242 CCTCGGTGATCGCGCGGTGTTCGCCGA.....TCGTGGGTG 279
   | | | | | | | | | | | | | | | | | |
131 YLeuGlyGlyValGlyGly.ValProGlyGlyValGlyValGlyVal 147
   | | | | | | | | | | | | | | | | | |
280 ACGGTGGCGCGCTCGGTGGGT....CCGGAGCGATGGCGCGCGGTTC 326
   | | | | | | | | | | | | | | | | | |
148 ProGlyGlyValGlyValGlyValProGlyGlyValGlyValGlyGl 164
   | | | | | | | | | | | | | | | | | |
327 GCAATCCGCGCGC.....TCCACCA 346
   : | | | | | | | | | | | | | | | | |
164 YValProGlyGlyValGlyGlyIleGlyGlyLeuGlyLeuGlyValS 181
   | | | | | | | | | | | | | | | | | |
347 GCCCGGTGTGTGTCGCGCGCG 366
   | | | | | | | | | | | | | | | | | |
181 erThrGlyAlaValValPro 187

```

seq_name: SwissProt_40:5E5_RAT

seq_documentation_block:

```

ID 5E5_RAT STANDARD; PRT; 825 AA.
AC Q63003.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 5E5 antigen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Brain;
RX MEDLINE=96015159; PubMed=8537300;
RA Suzuki E., Kojima N., Yoshimura K., Uyemura K., Akagawa K.;
RT "Cloning and sequence analysis of cDNA for a possible DNA-binding
RT protein 5E5 in the nervous system.";
RL J. Biochem. 118:122-128(1995).
CC -!- FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN NEURONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

CC	EMBL; D37934; BAA07153.1; -;
DR	DNA-binding; Nuclear protein; Antigen.
KW	SEQUENCE 825 AA; 86831 NW; AF667FE2FD555BDF CRC64;
SD	

alignment_scores:		
Quality:	126.50	Length: 183
Ratio:	1.265	Gaps: 8
Percent Similarity:	54.645	Percent Identity: 32.240

alignment_block:

US-09-462-480-2 x 5E5_RAT

Align seq 1/1 to: 5E5_RAT from: 1 to: 825

24 CAGCCAGGTGGG.....CGGCACCGCGGGCGCAACCAGCCGACGAGG 67
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
478 GlnAlaGlyGlyProArgGlyArgAlaArgGlyProArgGlnGlnAla 494
68 AAGCCCGCAGATGGCTCTCGGCACCATCGCTGTCGAACCATCCG 117
: ||| ||||| ||||| : : : : : ||| ||| ||| ||| ||| |||
494 argArgArgHisGlyProGlnArgArgGly.....Pro.Pro.Pro 507
118 CTGGCTGGTGATCAGCGCCGCGCGCGCGCGCTGCTG..... 159
||||| : : : : : ||||| : : : : : ||||| : : : : :
508 GlnAlaGlyGluGlyProGlyAspAlaThrLeuValLeuGlyLeuGl 524
160CGGCCGCGTCTACTCTGGCG 181
524 yThrThrSerGlyGluGlnArgAlaAspGlnSerGlnThrLeuProAlaLa 541
182 CAGGTGGGTGGTAGCCCG..CACGCCGCTGATCTCAGCTGATCGAAAA 230
: : ||||| : ||| ||||| ||| : : : : : ||||| : : : : :
541 euAlaGlyAlaProThrAlaHisAlaHisAlaValProGlyProGlyPro 557
231 GGCGTTGCCCTCTGGTGATGCCGC.....GGCTGTTGCCGATCGT 274
||| : : : : : ||||| ||| ||| ||| ||| ||| ||| ||| |||
558 AlaAlaAlaThrLeuGlyGlyArgGlyArgArgGlySerTrpArg..... 572
275 CGGTGAGCTGGCGCGCTCCGGTGGTCCGGGAGCGATGGCCAGCG 324
573GlyGlyArgGlyGlyGlyAlaGlyAlaSerGlyGlyGlyA 587
325 TCGCATCCGG.....CGGCTCCACGACGCCCGCTCTGGTCCGCGC 365
: : ||| ||| : : : : : ||||| ||| : : : : :
587 rgGlyGlyArgGlyArgGlyGlyGlyArgGlySerGlyLeuSer 603
366 GGCACCGCTCC...GCAGGACGCTGAAGAAGACGACGAGCAGCTGGG 412
||||| : : : : : ||||| : : : : : ||||| : : : : :
604 GlyThrGluAspAlaGlySerProSerAlaArgGlyGluGlnAr 620
413 ACGAAGAGGACGACTGCTGAGCTCCCGTAATGACAACAGACTTCCC 462
||||| :
620 gArgArgGlyHisGlyProProAlaAlaGlyAlaAlaGlnValSerThra 637
463 ACCCGGGCCGAACACTTGCCAACTTTTGGCAGGAAGGTAAGAG 509
: : ||| ||||| : : : : : ||||| ||| : : : : : : : : : : :
637 rgGlyArgArgAlaArgGlyGlnArgThrGlyGluAlaGlnAsp 652

seq_name: SwissProt 40:KLF2 HUMAN

seq_documentation block:

seq_documentation_block:	
ID	STANDARD;
KLIF2_HUMAN	PRT; 355 AA;

AC Q9Y5W3; Q9UKR6; Q9UJS5;
STANDARD, FBI, 300 AM.

DT 30-MAY-2000 (Rel. 39, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)
DT 01-MAR-2002 (Rel. 41, Last sequence update)

DE Kruppel-like factor 2 (Lung kruppel-like factor).

GN KLF2 OR LKLF.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RN
TISSUP=Lung;
RC
MEDLINE=99231781; PubMed=10217429;
RX
Kozyrev S.V., Hansen L.L., Poltarau A.B., Domninsky D.A.,
RA
Kisselev L.L.;
RA
"Structure of the human CpG-island-containing lung Kruppel-like factor
RT (KLF) gene and its location in chromosome 19p13.11-13 locus.;"
RT
FEBS Lett. 448:149-152(1999).
RN
[2]
SEQUENCE FROM N.A.
RN
TISSUP=Lung;
RC
MEDLINE=99389728; PubMed=10458913;
RX
Wani M.A., Konkright M.D., Jeffries S., Hughes M.J., Lingrel J.B.;
RA
"cDNA isolation, genomic structure, regulation, and chromosomal
RT localization of human lung kruppel-like factor.;"
RT
Genomics 60:78-86(1999).
RN
[3]
SEQUENCE FROM N.A.
RN
Lee H.J., Kim M.K., Kim Y.H., Seo J.M., Lee H.M., Chung H.J.,
RA
Sohn M.Y., Hwang S.Y., Im S.O., Jung E.J., Kim J.C.;"
RA
"A catalogue of genes in the human dermal papilla cells as identified
RT by expressed sequence tags.;"
RT
Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RL
-!- FUNCTION: BINDS TO THE CACC BOX IN THE BETA-GLOBIN GENE PROMOTER
CC
AND ACTIVATES TRANSCRIPTION (BY SIMILARITY).
CC
-!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC
-!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC
FINGER PROTEINS.
CC

CC
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC

EMBL;	AF123344;	AAD25076.1;	-;
EMBL;	AF134053;	AAD55891.1;	-;
EMBL;	AF205849;	AAF13295.1;	-;
HSSP;	P08047;	1SP2.	
MM;	602016;	-;	
InterPro;	IPR000822;	Znf-C2H2.	
Pfam;	PF000096;	Zf-C2H2;	3.
SMART;	SM00355;	ZnF_C2H2;	3.
PROSITE;	PS00028;	ZINC_FINGER_C2H2_1;	3.
PROSITE;	PS00157;	ZINC_FINGER_C2H2_2;	3.
Transcription regulation;	Activator;	Zinc-finger; Metal-binding;	
DNA-binding;	Nuclear protein;	Repeat.	
DOMAIN	62	71	POLY-PRO.
DOMAIN	130	135	POLY-GLY.
DOMAIN	167	171	POLY-PRO.
DOMAIN	225	231	POLY-ALA.
DOMAIN	272	354	ZINC FINGERS.
2N_FING	272	296	C2H2-TYPE.
2N_FING	302	326	C2H2-TYPE.
2N_FING	332	354	C2H2-TYPE.
CONFLICT	43	43	S -> N (IN REF. 2).
CONFLICT	104	104	L -> P (IN REF. 1).
CONFLICT	175	175	P -> S (IN REF. 2).
CONFLICT	184	184	L -> M (IN REF. 2).
SEQUENCE	355	AA: 37419 MW: D5849C831D676AE1	CRC64;

```

alignment_scores:
  Quality: 125.00      Length: 217
  Ratio: 1.289         Gaps: 13
  Percent Similarity: 44.700      Percent Identity: 29.032

alignment_block:
  US-09-462-480-2/rev x KLF2 HUMAN

```

```
Align seg 1/1 to: KLF2_HUMAN from: 1 to: 355
523 CTGGACTACTTCTCTCTTTA..... 503
45 LeuAspPheIleLeuSerMetGlyLeuAspGlyLeuGlyAlaGluAla 61
502 .....CCTTCTCCGCAAAATGTTGGCAAGTCTTCCGCG 469
61 aProGluProProProProProAlaPheTyrTyrProGluP 78
468 CGGGTGGCGG.....GGAAGTCTGTGTCA... 443
78 roGlyAlaProProProTyrSerAlaProAlaGlyLeuValSerGlu 94
442 ...TTACGGGAGCTCACGAGTCTCTCTCTCCAGT...CGNCCCG 399
95 LeuLeuArgProGluLeuAlaProLeuGlyProAlaLeuHisGlyAla 111
398 TCGTCTTCTTCCAGCTCTCTGCGGAGCGGTGCGGCGGACCAAGA... 354
111 gPheLeuLeuAlaProProGlyArgLeuValLysAlaGluProGluAla 128
353 .....CCGGGCTGTGTAGCGCGCG 333
128 laAspGlyGlyGlyGlyTyrGlyCysAlaProGlyLeuThrArgGlyPro 144
332 .....GATTGCGAACCTGGCGCCATCGCTCC... 306
145 ArgGlyLeuLysArgGlyAlaProGlyProAlaAlaSerCysMetAr 161
305 .GGACCCACCGGAGCGGCGCCACCGTCACGAGCATCGGCAACAGCGG 257
161 glyProGlyGlyArgProProProProAspThrProProLeuSer. 177
256 CGGGATCACCGAGGGGCAACCGCTTTTCATCAGCTGACATCAGC 207
178 .....ProAspGlyProAlaArgLeu.....ProAlaPro 187
206 GCGTGGCGGTCAACGACCGCTCGCGCAGGTAGCGACTCGCGCGCGAG 157
188 GlyProArgAlaSerPheProPro...ProPheGlyGlyProGlyPheG 203
156 CAGGCGCGCGCGCGCGTGTG.....GGGCTGATCCACGAGCCAGCG 116
203 yAlaProGlyProGlyLeuHisTyrAlaProProAlaProAlaPheG 220
115 GATGCTTCGACGCGGAGTGTGTCGCGAGCAGGCCCATCTGCGCGCTTCC 66
220 yLeuPheAspAspAla.....AlaAlaAlaAla 229
65 TCGTGGCGGTGCGCGCGCGG.....GTGCGCGCC 34
230 AlaAlaLeuGlyLeuAlaProProAlaAlaArgGlyLeuLeuThrProPr 246
33 C 33
246 o 246
seq_name: SwissProt_40:HM1D_DROAN
seq_documentation_block:
ID HM1D_DROAN STANDARD; PRT; 606 AA.
AC P22544;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Homeobox protein OM(1D).
GN OM(1D).
OS Drosophila ananassae (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
```

514 TTTCCTCTCTTTACCTTCTCTCGCCAAATGTTGGCAGCTCTTCGCGCCCGG 465
 337 TyrLeuProLeuProSerPro.....IleTyrSerProPr 349
 464 GTGCGCGGGGAAGTCTGTGTTCATTACGGGAGCTCACCGAGTCGCTCTTCT 415
 349 opPro.....ValTyrSerProProPr 358

414 GTCCCAGTCGTCCTCGTCGTCTTCTTCACGGTCCTGGCGCGAGCGGTGCCG 365

358 roProSer.....TyrSerPro..... 363

364 GCGGACCAAGACCGGGCTGGTGGAGCCGCCGGATTGCGAACCTTGGCC 315

364ProProThrTyr.....LeuProProProPr 374

314 ATCGCTCCCGGACCCACCGGAGCGGGGCCACCCGTCACCGACGATCCGGC 265

374 oSerSerProProProSerPheSerProProThrProThrTyrrGluG 391

264 AACAGCCGCCGCATCACCGAGGGGCAACCGGCTTTTCGATCAGCTGAG 215

```

||:::|||||
391 lnsrprprprprprprAlaTyrSerprprprprAlaproPro 407

```

214 ACATCAGCGGCGTGCGGTCAACGAC..... 189

408 ThrTyrSerProProProProThrTyrSerProProThrTyrAl 424

THE UNIVERSITY OF CHICAGO

188CCACCTGCGCCAGGTAGCGACTCCGCGCGCAGCAGGCCCGCGC 146

424 aClnProProLeuPro
ProThrTyrSerProProProp
::: ||| |||

[illegible]

145 CCGCGCTGGGGCCTGATCCACCAGCCAGCGGATGGTTCGACAGCGGACTG 96

438 raa l a t y r s e r p r o p r o b p r o b t h r
:::
:::
:::
:::
:::
Tyr 448

[illegible]

95 GTGCCGAGCAGGCCCATCTGCGCGGCTTCCCTCGTCGGC ..GGGTTGCC 49

[illegible]

449 SEIPFOIOFIOMIUYISEIPFOIOFIOMIUYIAAGIUF 403

48 GCCGCCGGTGCCGCCACCTGG 27

4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 5

463 ОРОРОРОРОРОТОРНТУР 4 / 2

seq_name: SwisssProt_40:TPM5_DROME

```
seq_documentation_block:
ID    TPM5 DROME    STANDARD:
PRT: 504 AA
```

AC P49456;

DT 01-FEB-1996 (Rel. 33, Created)

DT	UI-FEB-1996 (Rel. 33, Last sequence update)	DT	30-MAY-2000 (Rel. 39, Last annotation update)
DT	30-MAY-2000 (Rel. 39, Last annotation update)	DT	30-MAY-2000 (Rel. 39, Last annotation update)

DE Tropomyosin 1, fusion protein 34.

GN TM1 OR TMII.

OS *Drosophila melanogaster* (Fruit fly).

OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insec

OC Eukariyota; Metazoa; Attilopoda; Tracheata; Insecta; Diptera; Brachycera; M
OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; M

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;
DN [1]

RN [I]
RP SEQUENCE FROM N.A.

RX MEDLINE=89127197; PubMed=2851721;

RA Hanke P.D., Storti R.V.;

RT⁺ The *Drosophila* melanogaster tropomyosin II gene produces a

[illegible]

RA Lynch E.D., Lee M.K., Morrow J.E., Welch P.L., Leon P.E., King M.-C.;
RT "Nonsyndromic deafness DFNA1 associated with mutation of a human
RL homolog of the Drosophila gene diaphanous.";
RN Science 278:1315-1318(1997).
[2]
RN SEQUENCE OF 218-817 FROM N.A.
RC TISSUE=Ovarian carcinoma;
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Watanabe M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 727-765 AND 1121-1145.
RC TISSUE=platelet;
RX MEDLINE=9525215; PubMed=7737110;
RA Reinhard M., Giehl K., Abel K., Haffner C., Jarchau T., Hoppe V.,
RA Jockusch B.M., Walter U.;
RT "The proline-rich focal adhesion and microfilament protein VASP is a
RL ligand for profilins.";
RN EMBO J. 14:1583-1589(1995).
CC -1- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
CC SERUM RESPONSE FACTOR. DPR PROTEINS COUPLE RHO AND SRC TYROSINE
CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY
CC SIMILARITY). IN HEARING IT MAY PLAY A ROLE IN THE REGULATION OF
CC ACTIN POLYMERIZATION IN HAIR CELLS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART, PLACENTA, LUNG,
CC KIDNEY, PANCREAS, LIVER, SKELETAL MUSCLE, AND COCHLEA.
CC -1- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
CC RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION
CC (BY SIMILARITY).
CC -1- DISEASE: DEFECTS IN DIAPH1 ARE A CAUSE OF AUTOSOMAL DOMINANT
CC NONSYNDROMIC SENSORINEURAL DEAFNESS 1 (DFNA1).
CC -1- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 1 (FH1) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 2 (FH2) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 3 (FH3) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).
CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLGY FAMILY. DIAPHANOUS
CC SUBFAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THAT IT SEEMS
CC TO INCLUDE INTRONIC SEQUENCE.
CC -1- DATABASE: NAME-Hereditary hearing loss homepage;
CC NOTE=Gene page;
CC WWW="http://www.uia.ac.be/dnalab/hhh/hhhgenes.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF051782; AAC05373.1; ..
DR EMBL: AK023345; BAB14533.1; ALU_SEQ.
DR MIM: 602121; ..
DR MIM: 124900; ..
DR InterPro: IPR003104; FH2.
DR Pfam: PF02181; FH2: 1.
DR SMART: SM00498; FH2: 1.
KW Coiled coil; Repeat; Deafness.
FT DOMAIN 63 260 GBD.
FT DOMAIN 157 457 FH3.
FT DOMAIN 460 563 COILED COIL (POTENTIAL).
FT DOMAIN 588 743 FH1 (PRO-RICH).

FT DOMAIN 748 1190 FH2.
FT DOMAIN 1015 1172 COILED COIL (POTENTIAL).
FT DOMAIN 1173 1187 DAD.
FT DOMAIN 1189 1192 ARG/LYS-RICH (BASIC).
FT CONFLICT 804 804 T -> TSKA (IN REF. 2).
FT CONFLICT 1132 1133 RK -> AE (IN REF. 3).
SQ SEQUENCE 1248 AA; 138978 MW; EDIF5147CF9A886 CRC64;

alignment_scores:
Quality: 120.50 Length: 149
Ratio: 1.826 Caps: 8
Percent Similarity: 44.295 Percent Identity: 33.557

alignment_block:
US-09-462-480-2/rev x DIAL_HUMAN ..

Align seg 1/1 to: DIAL_HUMAN from: 1 to: 1248

448 TGTGATTACGGAGCTCACCAGTCGTCTCTTCGTCCTCCAGTCGTCCTCG 399
||||| :||| ||| |||||
556 LeuSerAlaAlaAlaIleThrValProSerValProSerArg..... 570
398 TCGTCTTCTTCCAGCTCTCGCGGAGCGGTCCGCGCGGACGACGACCGG 349
||| :| ||||| ||| |||||
571AlaProValProAlaProProLeuProG 581
348 GCTGTGGAGCGCGGGATTGCGAACCTCGCCCATCGCTCCCGGACCCA 299
|| :||| ||| ||| |||||
581 ly.....AspSerGlyThrIleIleProProPro 590
298 CCG.....GAGCGCGCCACCGGTCAACGCGATCCGCGCAACAGCC 258
||| :||| ||| ||| |||||
591 ProAlaProGlyAspSerThrThrProProProProProProProPr 607
257 GCGGATCACCAG.....GGGCAACCGGCTTTTCGATCAGCTGAGACA 212
||||| ||| |||||:|||||
607 oProProProProLeuProGlyThrAlaIleSerProProProProL 624
211 TCAGCGCGGTCCGCGTCAACGACCCACCTCGCGCA..... 177
:||||| ||| ||||| |||||
624 euSerGlyAspAlaThrIleProProProProProProGluGlyVal 640
176 GGTAGCGACTCCGCGCGCAGCAGGCCC...CGCGCGCGCTGGGCGCTGA 130
||| ||| ||| ||| :||| |||||
641 GlyIleProSerProSerSerLeuProGlyThrAlaIleProProPr 657
129 TCCACGACCGCAGCGATGTTTCGACGCGGACTGTCGCGGACGACCCA 80
||||| ||| ||||| :||| |||||
657 oProProLeuProGly.....SerAlaArgIleProProProPro. 670
79 TCTCGCGCGGTCTCTCGTCGCGGTTCGCGCGCGGTCGCGCGCC 33
:||||| |||||:||||| |||||
671 ..ProProLeuProGlySerAlaGlyIleProProProProProPro 685

seq_name: SwissProt_40.VNUA_PVKA
seq_documentation_block:
ID VNUA_PVKA STANDARD; PRT; 1733 AA.
AC P33485;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Probable nuclear antigen.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021039; PubMed=2171211;
RA Vlcek C., Kozmik Z., Paces V., Schirm S., Schwytzer M.;
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely

oriented open reading frame: characterization of their promoter and enhancer regions.";
Virology 179:365-377(1990).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M34651; AAA47471.1;
PIR; B45344; B45344.
FT DOMAIN 112 117 POLY-THR.
FT DOMAIN 179 1733 GLY-RICH.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 271 298 POLY-PRO.
FT DOMAIN 304 308 POLY-ARG.
FT DOMAIN 883 889 POLY-GLY.
FT DOMAIN 1398 1405 POLY-GLY.
SEQUENCE 1733 AA; 172166 MW; 0C8CD8BB475BB5E2 CRC64;

alignment_scores:

Quality:	119.50	Length:	231
Ratio:	1.160	Gaps:	12
Percent Similarity:	44.589	Percent Identity:	28.571

alignment_block:

US-09-462-480-2 x VNUA_PRVKA ..

Align seg 1/1 to: VNUA_PRVKA from: 1 to: 1733

31	GTGGCGCGCACCGG	GGCGG	50
840	ValGlyGlyGluGlyArgLeuGlyGlyProArgValGlyLeuAlaG	AlaG	856
51	CAACCCAGCCGACGAGGAGCGGCGGCTGCTCGGCGCACGATC	100	
856	YargAspAlaAlaGluAlaValGlyArgGlyValLeuGlyHisGlyP	873	
101	CG....CTGTGCAACATCCGCTG.....GCTGTGGATCAGCG	135	
873	roGluArgAlaProGluProValValLeuGlyGlyGlyGlyGly	889	
136	CCCAGCGCGCGCGGCGCTG.....	156	
890	GlnGlnArgGlySerglyValArgSerglyProGluSerglyAlaAl	906	
157	C	157
906	aLeuAlaProGlyProProValLeuPheValValAlaValAlaVal	923	
158	TGCGCGCGGAG.....TCGCTACTC	177	
923	alProAlaGluGlyArgAlaGlyGluProLeuValLeuAlaValPro	939	
178	GGCGAGGTGGTGTGTACCCGCGGCTGTGTCTCAGCTCATCGA	227	
940	GlyAlaAlaGly.....ProGlyArgAlaAlaLeuLeuLe	951	
228	AAAGCCGGTTCGCCCTCGGTGATGCCGCGCTGTTCGCCGATCGCGG	277	
951	uAlaProLeuGlyArgTrpValArg....AlaGlyGlyGlyAlaGlyV	967	
278	TGACGGGTGGCGCGCTCGGTGGGTCCGGGAGCG	312	
967	alAlaGlyGlyAlaGlyGluAlaGlyLeuGlyAlaGlyAlaGlyLeuGly	983	
313	ATGGCCAGAGGTTCGAATCCGGCGGCTCCAC.CAGCCCGGTCTGGTCG	361	
984	AlaGlyAlaGlyLeuGlyAlaGlyGlyAlaGlyGlyProGlyAlaGlyG	1000	

362	CGCGCGCACCGCTCGCGAGGAGCGTGAAGAAGACGA.....	398
1000	uAlaGly.....GlyGlyAlaArgArgArgArgArgTrpA	1014
399CGAGGACGACTGGGACGAGA	419
1014	spAspGluAlaGlyLeuLeuGlyProGluArgGlyGlnAlaGlyArgGly	1030
420	GGAGCGACTGGTGACTCCCGTAATGACACACAGACTCCC.....	GGCCA 463
1031	LeuArgGlyProGlyProArgGlyGlyLeuGlyGluProGlyArgGlyH	1047
464	CCCGCGCGGAGGAGACTTCCCAACATTTTGGCGAGGAGGTAATA	506
1047	sValGlyArg.....GlyGluGluGlyArg	1055

seq_name: SwissProt_40:EXLP_TOBAC

seq_documentation_block:

ID EXLP_TOBAC STANDARD; PRT; 426 AA.

AC Q03211;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pistil-specific extensin-like protein precursor (PELIP).
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, PETITE HAVANA; TISSUE=Pistil;
RX MEDLINE=93005740; PubMed=1392607;
RA Goldman M.H., Pezzotti M., Seurinck J., Mariani C.;
RT 'developmental expression of tobacco pistil-specific genes encoding novel extensin-like proteins.';
RL Plant Cell 4:1041-1051(1992).
CC -!- TISSUE SPECIFICITY: PISTIL (STIGMA AND STYLE TISSUE).
CC -!- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN FLORAL BUDS AFTER PISTIL DIFFERENTIATION AND LEVELS GRADUALLY INCREASE DURING FLOWER DEVELOPMENT TOWARD ANTHESIS. LEVELS GRADUALLY DECREASE AFTER POLLINATION AND ARE ABSENT BY THE SIXTH DAY AFTER POLLINATION.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; Z14019; CAA78397.1; -;
CC PIR; JQ1696; JQ1696.
DR InterPro; IPR000419; Pollen_Ole_e_1.
DR Pfam; PF01190; Pollen_Ole_e_1; 1.
KW Structural protein; Signal; Repeat; Glycoprotein.
FT SIGNAL 1 23
FT CHAIN 24 426 PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN.
FT DOMAIN 69 73 4 X 5 AA REPEATS OF S-P(4).
FT REPEAT 69 73 1.
FT REPEAT 76 80 2.
FT REPEAT 83 87 3.
FT REPEAT 178 182 4.
FT CARBOHYD 310 310 N-LINKED (GLCNAC...) (POTENTIAL).
FT SEQUENCE 426 AA; 44278 MW; 51A495CC94017812 CRC64;

alignment_scores:

Quality:	119.00	Length:	175
Ratio:	1.451	Gaps:	9

Percent Similarity: 46.857 Percent Identity: 28.571

alignment_block:

US-09-462-480-2/rev x EXLP_TOBAC ..

Align seg 1/1 to: EXLP_TOBAC from: 1 to: 426

```
524 GGTGACACTTCTCTCTTTACCTTCCTCGCAAAATGTTGGCAAGTCT 475
|||||
113 AlAGlyLeuProLeu.IleProAsnLeuProAspValProProIleG 129
474 TC.....CGGCGGGTGGCGCGGAAGTCTG 449
129 lYgLyglyProProValAsnGlnProLysProSerProSerProLeu 145
448 TTGTCTATTAGGGAGCTCACCAGTCTCTCTCTGTCCTCCAGTCGCTCCG 399
146 VallysProProProProSerProSerProCysLysProSerProProAs 162
398 TGTCTCTTCTCACGCTCTCGCGAGCGGTGCGCGCGGACGACACCGCG 349
162 pElnSerAlaLysGlnProGlnProProProAlaLys..... 175
348 GCTGTGGAGCGCGCGGATTGGGAACCTTGGCCCATCGCTCCGGACCCA 299
176 .....GlnProSerProPro...ProProPro 183
298 CGGAGCGCGCGCACCGCTCAGCAGCATCGCGCAACAGCGCGCGCATC 249
184 ProProValLysAlaProSerProSerProAlaLysGlnProProPro 200
248 ACCGAGGGGCAACCGGCTTTTCGATCAGCTGAGACATCAGCGCGTGGC 200
200 oPro.....ProProValLysAlaProSerProSerProAlaThrg 214
199 G.....GTCAACGACCCACTGCGCGCAGTAGGACTCGCGCGGAGC 156
214 lNProProThryLysGlnProProProPro.....ProArgAla 226
155 AGGCGCGCGCGCGCTGGGCGCTGATCCACGACGCGGATGTTTGA 106
227 LysLysSerProLeuLeuProProProProProValAla.....TyrPr 241
105 CAGCGGACTGTGGCGGACGAGCCCATCTGCGCGGCTTCCGTCGCGGTG 56
241 oProValMetThrProSer...ProSerProAlaAlaGluProProIleI 257
55 GGTTCGCGCGCGCGTGGCGGCC 33
257 leAlaProPheProSerProPro 264
```

seq_name: SwissProt_40:FMN_CHICK

seq_documentation_block:

```
ID FMN_CHICK STANDARD; PRT; 1213 AA.
AC Q05858;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Formin (Limb deformity protein).
GN LD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEHORN; TISSUE-Embryo;
RX MEDLINE=92112031; PubMed=1730407;
RA Trumm A., Blundell P.A., de la Pompa J.L., Zeller R.;
RT "The chicken limb deformity gene encodes nuclear proteins expressed
in specific cell types during morphogenesis.";
```

```
Genes Dev. 6:14-28(1992).
-1- FUNCTION: IS IMPORTANT FOR MORPHOGENESIS OF LIMB AND KIDNEY AND
MAY BE INVOLVED IN DETERMINING DORSOVENTRAL NEURAL TUBE POLARITY
AND MOTOR NEURON INDUCTION. IT MAY ALSO HAVE A FUNCTION IN
DIFFERENTIATED CELLS OR BE INVOLVED IN MAINTAINING SPECIFIC
DIFFERENTIATED STATES.
-1- SUBCELLULAR LOCATION: Nuclear.
-1- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE
PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION
IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE
TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.
-1- TISSUE SPECIFICITY: PRESENT IN THE ADULT BRAIN, KIDNEY,
BRAIN, HEART AND INTESTINE AND THROUGHOUT THE EMBRYO.
-1- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE PROTEIN IS
EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESENCHYMAL
COMPARTMENT. PREDOMINANTLY IN THE POSTERIOR REGION DURING
KIDNEY MORPHOGENESIS, EXPRESSION IS INITIALLY RESTRICTED TO
THE EPITHELIAL COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.
-1- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 1 (FH1) DOMAIN.
-1- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 2 (FH2) DOMAIN.
-1- SIMILARITY: BELONGS TO THE FORMIN HOMOLGY FAMILY. CAPPUCCINO
SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X62681; CAA44555.1; -.
DR InterPro; IPR003104; FH2.
DR InterPro; IPR001265; Formin.
DR Pfam; PF02181; FH2; 1.
DR PRINTS; PR00828; FORMIN.
DR SMART; SM00498; FH2; 1.
DR Nuclear protein; Developmental protein; Coiled coil;
KW Alternative splicing.
FT DOMAIN 428 450 COILED COIL (POTENTIAL).
FT DOMAIN 503 572 COILED COIL (POTENTIAL).
FT DOMAIN 652 751 FH1 (PRO-RICH).
FT DOMAIN 766 1171 FH2.
FT DOMAIN 1050 1125 COILED COIL (POTENTIAL).
SQ SEQUENCE 1213 AA; 135240 MW; ADE3EF0B3FB9D862 CRC64;

alignment_scores:
Quality: 118.00 Length: 107
Ratio: 2.511 Gaps: 5
Percent Similarity: 43.925 Percent Identity: 31.776

alignment_block:
US-09-462-480-2/rev x FMN_CHICK ..
Align seg 1/1 to: FMN_CHICK from: 1 to: 1213
353 CCCGGCTGTGGAGCGCGGATTCGGAACCTGGCCCATCGCTCCGCG 304
|||||
678 ProGlyLeuValProProPro.....ProLeuProThrGl 690
303 ACCACCGGAGCGCGCACCGTCACCGACGATCCGGCAACAGCGCGG 254
|||||
690 yProThrSerValThrProHisPheAlaPheGlyProProLeuProProG 707
253 GCATCAGCGGAGCGCAACCGCTTTTCGATCAGCTGAGACATCAGCGG 204
|||||
707 lNLeuSerGluGlyCysArgAspPhe..... 715
203 GTGCGGGTCAACGACCCACCTGCCAGGTAGCGACTCCGCGCGCAGCAG 154
|||||
716 ...GlnAlaProAlaProProAlaPro.....Pr 724
```

```
153 GCCCGCCGCGCGTGGGCTGATCCACGAGCGGATGTTTCGACA 104
    ||| |||:||||||| |||
724 oProLeuProGlyLeuGlyProValPro..... 734
103 CGCGACTGGTCCGAGCAGCGCCATCGCGGGTCTCTGCTGGCTGGG 54
    ||| ||| :|:|:|:|
735 .....ProProLeuPro.....GlySerGly 741
53 TTGCGCGCGCGGTGGCGGCC 33
||||| ||| ||| |||
742 LeuProProProProProPro 748
seq_name: SwissProt_40:HK61_RAT
```

```
seq_documentation_block:
ID HK61_RAT STANDARD; PRT; 365 AA.
AC O35762;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Homeobox protein NKX-6.1.
GN NKX6A OR NKX6.1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=pancreatic islets;
RA Jorgensen M.C., Madsen O.D., Serup P.;
RT "The cDNA sequence of rat NKX6.1.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY BE IMPORTANT FOR CONTROL OF ISLET DEVELOPMENT AND/OR
CC REGULATION OF INSULIN BIOSYNTHESIS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: PANCREATIC BETA CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF004431; RAB61665.1; -
CC HSP; P02833; 9ANT.
CC InterPro: IPR000047; HTH_repressr.
CC Pfam: PF001356; Homeobox.
CC PRINTS: PF00046; homeobox; 1.
CC PRINTS: PR00024; HOMEBOX.
CC SMART: SM00389; HOX; 1.
CC PROSITE: PS00027; HOMEBOX_1; 1.
CC PROSITE: PS50071; HOMEBOX_2; 1.
CC Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DOMAIN 49 61 POLY-SER.
FT DOMAIN 119 132 POLY-SER.
FT DOMAIN 136 151 POLY-ALA.
FT DOMAIN 169 174 POLY-PRO.
FT DNA_BIND 237 296 HOMEBOX.
FT DOMAIN 323 326 POLY-ASP.
SQ SEQUENCE 365 AA; 37689 MW; C4AAB702D051F1F2 CRC64;
```

```
alignment_scores:
Quality: 117.50 Length: 217
Ratio: 1.141 Gaps: 8
Percent Similarity: 47.465 Percent Identity: 26.728

alignment_block:
US-09-462-480-2/rev x HK61_RAT
```

```
Align seq 1/1 to: HK61_RAT from: 1 to: 365
512 TCCTCTTTACCTTCTCTCGCAAAATGTTGCAAGTCTTCGCGCCGGGT 463
    ||| ||| :|:|:|:| |||:|:|:|:|
13 SerAlaPheLeuLeuSerProProLeuAlaAlaLeuHisSerMetal 29
462 GCGCGGAAGTCTGTGCTATTACGGGAGCTCAC.....CAGTCGT 422
    : |||:|:| :|:|:|:|
29 aGluMetLysThrProLeuTyProAlaAlaTyProProLeuProThrG 46
421 CCTCTTCGTCCTCAGTCCTGCTGCTCTTCTTCACGCTCCTCGCGCAGC 372
    : |||:|:|:|:|:|:|:|:|:| ||| :|:|
46 LyProProSerSerSerSerSerSerSerSerSerSerSerPro 62
371 GGTGCGCGCGCACGACCGCGGCTGGTGAGCGCGCGGATTCGCAACC 322
    ||| ||| :|:|:|:| :|:|:|:| |||:|:|
63 ProLeuGlyAlaHisAsnProGlyGlyLeuLysProProAlaGlyG 79
321 CTGCGCCATCGCTCCGCGACCCACCGGAGCGG..... 290
    ||| ||| :|:|:|:|:|:|
79 LyLeuSerSerLeuGlySerProProGlnGlnLeuSerAlaAlaThrPro 95
289 .....CGCCACCGCTCACCGAGCATCCGCAAA... 263
96 HisGlyIleAsnAspIleLeuSerArgProSerMetProValAlaSerG 112
262 .....CAGCGCGCGCATCACGAGGGGCAACCGGCTTTTCGATCAG 220
    ||| ||| ||| ||| ||| ||| ||| ||| |||
112 yAlaAlaLeuProSerAlaSerProSerGlySerSerSerSerSers 129
219 CTGAGACATCAGCGCGTGGGGTCAACGACGCCACCTCGCGCAGGTAGCG 170
    :|:|:|:|:|:|:|:|:|:| :|:|:|:|:|:|
129 erSerAlaSerAlaThrSerAlaSerAlaAlaAlaAlaAlaAlaAla 145
169 ACTCGCGCGCAGCAGCGCGCGCGCGG..... 141
    :|:| |||:|:|:|:| |||
146 AlaAlaAlaAlaAlaAlaSerProAlaGlyLeuLeuAlaGlyLeuPr 162
140 .....CTGGGGCTGATCCACCGAGCGCGGATGTTTCGACA 104
    |||:|:|:|:| ||| ||| ||| :|:|
162 oArgPheSerSerLeuSerProProProProProProGlyLeuTyrrPheS 179
103 CGCGACTGGTCCGAGCAGCGCCATCTCGCGGGCTTCTCTCGCTGGCTGGG 54
    || ||| :|:|:|:|:|:|:|:| ||| :|:|:|:|:|
179 er.....ProSerAlaAlaAlaAlaAlaAlaValGly 189
53 .....TTGCGCGCGCGGTGGCGG..... 36
    ||| ||| :|:|:|:|
190 ArgTyProLysProLeuAlaGluLeuProGlyArgThrProIlePheTr 206
35 .....CCACCTGGCTGAACAACGACGTCACCTGC 6
206 pProGlyValMetGlnSerProProTrpArgAspAlaArgLeuAlaCys 222
```



```

251 TGCCGGGGGCTGTTCGGGATCGTCGGTGACGGGTGGCGCGCTCCGGTG 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 GGTCGGGAGGATGGCCAGGTTTCGAATCCGGCGGCTCCACAGCCC 350
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
351 GGGTCTGGTGGCGCGGACCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspGp 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 AGGACGACTGGGACGAGAGGACGACTGG 429
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq_name: sp_bacteriap:033085

```

seq_documentation_block:
ID O33085; PRELIMINARY; PRT; 302 AA.
AC O33085;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 32.1 KDA PROTEIN (PPE-FAMILY PROTEIN).
GN M00051 OR MLCB628.14C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Eiglmeyer K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jajelski K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; Y14967; CAA75201.1; -.
DR EMBL; AL583917; CAC29559.1; -.
DR Leproma; ML0051; -.
DR InterPro; IPR000030; PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 302 AA; 32135 MW; EDAC4CC0AF3BF3D0 CRC64;

```

```

alignment_scores:
Quality: 198.00 Length: 126
Ratio: 2.329 Gaps: 4
Percent Similarity: 67.460 Percent Identity: 41.270

```

alignment_block:

US-09-462-480-2 x O33085

Align seg 1/1 to: O33085 from: 1 to: 302

```

1 CTGACGAGGTGAGTGTTCAGCCAGGTGGCGGCGGCGGCGGCGG 50
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 MetGlnGlnValLysSerLeuPheThrSerIleAspSerThrGlyValty 206
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 CAACCCGCGCGAC.....GAGGAAGCCGCGCAGATGGCGCTGC 88

```

```

::: |||::: |||:|||||:|||||:|||||:|||||:
206 rThrSerAlaGlnArgGlyAspThrGluSerAlaHisArgIleGlyLeup 223
89 TCGGCACGAGTCGCTGTCGAACCATCGCTGGCTGGTGGATCAGGCCCC 138
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||
223 heGlyAlaSerThrLeuSerHisProLeuValGlyIleThrGlyThr 239
139 AGCGCGGCGCGGCGCTGCTCGCGCGGAGTCGCTACCTGGCGCAGGTGG 188
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||
240 ThrThrAspThrArgLeuLeuCysAlaGluSerLeuProSerAlaSerGl 256
189 GTCGTTGACCGCACGCGCTGATGTCT.....CAGCTGATCGAAAGC 232
|||||:|||||:|||||:|||||:|||||:|||||:|||||
256 ySerLeuAlaTrpThrProLeuMetThrGlnPheGlnLeuIleAspLys 273
233 CGGTTGCCCGCTCG.....GTGATCGCGCGGCTGCTGCC 267
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
273 erIleAlaProGluProArgGlnArgValMetLeuProTrpAlaAla 289
268 GGATCGTCGGTGACGGGTGGCGCGCTCCGCTGGTGGCGGAGCGATGG 317
|||||
290 GlySerPro.....G.....GL 293
318 CCAGGTTGCGCAATCCGGCGGCTCCACC 345
|:|||||:|||||:|||||:|||||
293 yHisAsnAlaGlnAspGlyGlyThrThr 302

```

seq_name: sp_bacteria:Q9EY22

```

seq_documentation_block:
ID Q9EY22; PRELIMINARY; PRT; 251 AA.
AC Q9EY22;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 28.0 KDA PROTEIN.
OS Paracoccus pantotrophus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Paracoccus.
OX NCBI_TaxID=82367;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB17;
RX MEDLINE=20566691; PubMed=11114924;
RA Bardischewsky F., Friedrich C.G.;
RT "Identification of ccdA in Paracoccus pantotrophus GB17: disruption of
RT ccdA causes complete deficiency in c-type cytochromes.";
RL J. Bacteriol. 183:257-263(2001).
DR EMBL; AF308446; AAG29834.1; -.
DR InterPro; IPR002965; P_Rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
KW Hypothetical protein.
SQ SEQUENCE 251 AA; 27997 MW; 0A1528E3539F2195 CRC64;

```

```

alignment_scores:
Quality: 145.50 Length: 158
Ratio: 1.914 Gaps: 9
Percent Similarity: 48.101 Percent Identity: 34.177

```

alignment_block:

US-09-462-480-2/rev x Q9EY22

Align seg 1/1 to: Q9EY22 from: 1 to: 251

```

463 TGGCGGGGAAGTCTGTGTTCATTACGGGAGCTCACAGTCGTCTCTTCG 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||
87 TrpProGlyProArgMet.....ArgProAsnAr 96
413 T...CCAGTCGTCGTCGTCTTCTTTCACGCTCTCTCGCGGCGGTCG 367
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
96 gMetProAsnArgProArgSerArgAlaAlaAlaProGlyArgAlaAla 113

```

```
366 CGGCG.....CGACGACCCGGGCTGGTGGAG...CGCGGGATT 329
...
113 laSerArgLysAsnArgProLysProTrpArgSerArgArgVal 129
...
328 CGAACCTGGC.....CCATCGCTCCCGGACCCACCGGA 294
...
130 SerLysProSerArgArgSerArgProSerSerLeuProArgProH 146
...
293 CGCGCGCACCCGACCGGATCCGCGACACACCGCGGCATCACCGA 244
...
146 sArgArgThrProArgArgHisArgMetArgArgProGlySer.... 161
...
243 GGGGGCAACGGCTTTCGATCAGTCAGCATCAGCGCGGCTGGGTCA 194
...
162 .....ArgLeuProSerLeuProSerAlaArgArgPro 172
...
193 ACGACCCACCTGCCGAGTAGCAGTCGCGGCGCAGCGCCGCGGCC 144
...
173 ThrProArgLeuLeuAsnProArgProThrProArgLysProArgPr 189
...
143 GCGCTGGGGCTGATCCACCGACCGGATGTTTCGACACGCGACTGGT 94
...
189 oArg...AsnProLysPro.....SerGlyGlnAr 198
...
93 GCCGAGCAGGCCCATC.....TGGCGGCTTCTCGTCGGCTG 56
...
198 gProLysArgProLeuHisProLysProCysArgArgProSerProSerP 215
...
55 GGTTCGCGCGCGGTCGCGGCC 33
...
215 roProProSerProLeuProPro 222

seq_name: sp_human:Q9BU37

seq_documentation_block:
ID Q9BU37 PRELIMINARY; PRT; 358 AA.
AC Q9BU37;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO WISKOTT-ALDRICH SYNDROME PROTEIN INTERACTING PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA.;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC002914; AA02914.1; -.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR003124; WH2.
DR Pfam; PF02205; WH2; 1.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR SMART; SM00246; WH2; 1.
SQ SEQUENCE 358 AA; 36464 MW; D008B60E60EE94EA CRC64;
```

```
alignment_scores:
  Quality: 143.50      Length: 187
  Ratio: 1.966
Percent Similarity: 39.037 Percent Identity: 29.947

alignment_block:
US-09-462-480-2/rev x Q9BU37 ..
Align seg 1/1 to: Q9BU37 from: 1 to: 358

508 TCTTTACTCTCCGCAAAATGTGGCAAGTCTCCGGCCCGGGTGCC 459
...
142 SerProProSerGlyProGlyArgPheProValProSerProGlyHisAr 158
```

```
458 GGAAGTCTGTGTGTCATTACGGAGCTCACGAGTCGTCTCTTCGTCCCA 409
...
158 gSerGly.....Prop 162
...
408 GTCGTCTCTGTCGTCTTTCACGCTCTCGCGGAGCGGTGCCGCGCGGA 359
...
162 roGluProGlnArg.....AsnArgMetProProArg 173
...
358 CCACAGCCGGGCTGGTGAGCGCGGATTCGAACCTGCCCATCGCT 309
...
174 ProAspValGlySerLysProAspSerIleProProValProSerTh 190
...
308 CCGGACCA..... 299
...
190 rProArgProIleGlnSerSerLeuHisAsnArgLysSerProProValP 207
...
298 .....CGGAGCGCGCCACCGTCACCGATCCGCGACGATCCGCGACGCG 257
...
207 roGlyGlyProArgGlnProSerProGlyProThrProProProPro 223
...
256 CCGCATCACCGAGGGCAA.....CCGCGC 231
...
224 ValArgAspProProGlyArgSerGlyProLeuProProProProPro 240
...
230 TTTTCGATCAGTCAGATCAGCGCGGTGGGTCAACACGCCACCTGC 181
...
240 lSerArgAsnGlySerThrSerArgAlaLeuProAlaThr..ProGlnLe 256
...
180 GCCA.....GGTAGCGACTCGCGCGCAGCAGCGCGCGCGCGCGC 140
...
256 uProSerArgSerGlyValAspSerProArgSerGlyProArgProProL 273
...
139 TGGGCGCTGATCCACCGACCGGATGTTTCGACAGCGGACTGGTGCGG 90
...
273 euProProAspArgPro..... 278
...
89 AGCAGGCCCATCTGCGCGGCTTCTCGTCGGTGGTTCGCGCGCGCGGT 40
...
279 .....SerAlaGlyAlaProProProPr 286
...
39 GCCGCGCCACC 30
...
286 oProProSer 289

seq_name: sp_bacteria:Q9XDH2

seq_documentation_block:
ID Q9XDH2 PRELIMINARY; PRT; 763 AA.
AC Q9XDH2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=20090472; PubMed=10627046;
RA Espitia C., Lacleite J.P., Mondragon-Palomino M., Amador A.,
RA Campuzano J., Martens A., Singh M., Cicero R., Zhang Y., Moreno C.;
RT "The PE-PGRS glycine-rich proteins of Mycobacterium tuberculosis: a
RL new family of fibronectin-binding proteins?";
RL Microbiology 145:3487-3495(1999).
DR EMBL; AF071081; AAD41594.1; -.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistil_extensin.
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01218; PSTLEXTENSIN.
```



```

175 .....SerLeuSerProAspGlyPro 181
212 .....ATCAGCGCGTGC GGGTCAACGACCCACCTGC GCCCAGG 175
182 AlaArgMetProAlaProGlyProArgAlaSerPheProPro....ProPh 197
174 TAGCGACTCCGCCCGCAGCAGCGCCGCCGCCGCTG.....GGCG 134
197 eGlyGlyProGlyPheGlyAlaProGlyProGlyLeuHisTyAlaProp 214
133 CTGATCCACCGCAGCGGATGTTCCACACCGCAGCTGTCGCCGAGCAGG 84
214 roAlaProProAlaPheGlyLeuPheAspAla..... 225
83 CCCATCTCGCGCGCTCTCTCGTCGGCTGGGTGGTCGCCGCCGCG..... 42
226 .....AlaAlaAlaAlaAlaLeuGlyLeuAlaProProAlaAlaAr 240
41 .....GTGCCGCC 33
240 gGlyLeuLeuThrProPro 246
seq_name: sp_mammal:Q9N1P0

seq_documentation_block:
ID Q9N1P0 PRELIMINARY; PRT; 818 AA.
AC Q9N1P0;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE SUBMAXILLARY MUCIN (FRAGMENT).
GN BSM1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutele
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; B
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRED ANGUS;
RX MEDLINE=20223253; PubMed=10759843;
RA Jiang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.
RT "The central domain of bovine submaxillary mucin consists
RT tandem repeats of 329 amino acids: chromosomal localization
RT BSM1 gene and relations to ovine and porcine counterparts.
RL Eur. J. Biochem. 267:2208-2217(2000).
DR EMBL; AF178428; AAF67279.1; -.
FT NON_TER 1
FT NON_TER 818
FT NON_TER 818
SQ SEQUENCE 818 AA; 73014 MW; B5B84F84F66F86B CRC64;

alignment_scores:
Quality: 128.50 Length: 182
Ratio: 1.397 Gaps: 9
Percent Similarity: 50.549 Percent Identity: 30.220

alignment_block:
US-09-462-480-2/rev x Q9N1P0 ..
Align seg 1/1 to: Q9N1P0 from: 1 to: 818

464 GTGGCGGGGAAGTCTGTTGTCAATACGGGAGCTCACAGTCGTCCTTC 415
355 lLeSerGlyThrAsnValProValSerGlyAlaProValThrProGlySe 371
414 GTCCAGTCGTCCTCG.....T 398
371 rSerAlaGlySerSerGlyAlaProGlyThrGlyGlyProGlySerGluT 388
397 CGTCTTCTTCACTCTCGCGCAGCGTGC CGCGCGCAGCACGACCCGCG 348

```

37 GGCACCGCGCGGGCAACCCAGCCAGCAGAACCCCGCAGATGGCGCT 86
||| |||||:|||||:|:| |||
1146 GlyProGlyGlyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyPro 1162
87 GCTGGCACCAGTCCGCTGTGCGAACCATCCGCTGGCTGGGTGATCAGGCC 136
|||:|||||:|:| |||||:|||||
1162 cGlyGlyAlaGlyPro.....GlyGlyAlaGlyPro 1172
137 CCAGCGCGGGCGCGGGCTGCTGTCGCGCGAGTGTCTACCTGCCGCAAGT 186
||:|||||:||||| |||:|:|
1172 roGlyGlyAlaGlyGlyAlaGlyGlyAlaGlyGlySerGlyGlyAlaGly 1188
187 GGGTCTGTGACCCGACGCGCGCTGATGTCACGCTG..... 222
||| ||| |||||:|||||
1189 GlySer...GlyGlyThrThrIleIleGluAspLeuAspIleThrIleAs 1204
223 ...ATCGAAAGCGGTGTGCCCTCG...GTGATCGCGCGCGCTGTG 265
:|: |||||:|:| |||:|:|:|
1204 pGlyAlaAspGlyProIleThrIleSerGluGluLeuProIleSerGlyA 1221
266 CCGATGCTCGGTGACGGGTGGCGCGCTCCG.....GTGGTTCGGGA 309
|||:|:| |||||:|:| |||||:|:|
1221 laGlyGlySerGlyProGlyGlyAlaGlyProGlyGlyValGlyProGly 1237
310 CGATGGCCAGGTCGCAATCCGGCGGTCCACACGCGCGGTCTGGT 359
:|: |||||:|:| |||||:|:| |||||:|:|
1238 GlySerGlyProGlyGlyValGlyProGlyGlySerGlyProGlyGlyVal 1254
360 CGCGCGC 366
1254 lGlyPro 1256

seq_name: sp_invertebrate:044358

```

seq_documentation_block:
ID      044358      PRELIMINARY;      PRT;      871 AA.
AC      044358;
DT      01-JUN-1998 (TremBLrel. 06, Created)
DT      01-JUN-1998 (TremBLrel. 06, Last sequence update)
DT      01-JUN-2001 (TremBLrel. 17, Last annotation update)
DT      01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE      FLAGELLIFORM SILK PROTEIN (FRAGMENT).
GN      FLAG.
OS      Nephila clavipes (Orb spider).
OC      Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC      Araneomorphae; Entelegynae; Araneioidea; Tetragnathidae; Nephila.
OX      NCBI_TaxID=6915;
RN      [1]
RS      SEQUENCE FROM N.A.
RX      MEDLINE=98153262; PubMed=9480768;
RA      Hayashi C.Y., Lewis R.V.;
RT      "Evidence from flagelliform silk cDNA for the structural basis of
RT      elasticity and modular nature of spider silks.";
RL      J. Mol. Biol. 275:773-784(1998) .
DR      EMBL; AF027972; AAC38846.1; -.
DR      InterPro; IPR000087; Collagen.
DR      InterPro; IPR000209; Peptidase_S8.
DR      PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR      NON_TER      871      871
SQ      SEQUENCE      871 AA; 1FAJE3B7E0C5983A CRC64;

```

```

alignment_scores:
  Quality: 128.00      Length: 132
  Ratio: 1.707         Gaps: 6
  Percent Similarity: 56.818      Percent Identity: 35.606

alignment_block:
  US-09-462-480-2 x 04358      ..

Align seg 1/1 to: 04358 from: 1 to: 871

```

37 GGCACCGGGCGGGCAACCCAGCCGACGAGGAAGCCGGCAGATGGGGCT 86

```

377 GlyProGlyGlyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyPr 393
87 GCTCGGCACCAAGTCGCTGTCGACCAATCCGCTGGCTGGTGGATCAGGCC 136
393 oGlyGlyTyGlyProGlyGlySerGlyPro.....GlyGlyAlaGlyP 408
137 CCAGCGGGGGCGGGCGCTGCTGCGCGGAGTGCCTACTGCGCGCAGGT 186
408 roser....GlyAlaGlyLeuGlyAlaGly.....ProGlyGlyAla 421
187 GGGTCGTGACCCGACCGCGCTGATGCTCAGCTGATCGAAAGCCGGT 236
422 Gly..... 422
237 TGCCCCCTCGGTGATCGCGGGCTGTTGCGGATCGTGGTGACGGGTG 286
423 .....LeuGlyGlyAlaGlyProGlyGlyAlaGlyThrSerG 435
287 GCGCGGCTCCG.....GTGGTCCGGAGCGATGGCGCGGTCGCAA 330
435 lyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyGlnGlyAspAla 451
331 TCCGGCGGTCACCAAGCC....GGTCTGCTGCGCGCGCACCGCTCG 376
452 GlyProGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyValGlyAr 468
377 GCGAGGAGCGTGAAGACAGCAGCAGGACGAGTGGACGAGAGGA 422
468 gGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyAlaGlyArgGly 483
```

seq_name: sp_rodent:Q9WUE8

```

seq_documentation_block:
ID Q9WUE8 PRELIMINARY; PRT; 2087 AA.
AC Q9WUE8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SHANK1A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=99360650; PubMed=10433268;
RA Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtschanoff J.,
RA Weinberg R.J., Worley P.F., Sheng M.;
RT "Shank, a novel family of postsynaptic density proteins that binds to
RT the NMDA receptor/PSD-95/GKAP complex and cortactin.";
RL Neuron 23:569-582(1999).
DR EMBL; AF131951; AAD29417.1; -.
DR HSSP; P00519; IABL.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR002965; P-rich_extensn.
DR InterPro; IPR001680; SAM.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00536; SAM; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50002; SH3; 1.
```

```

KW ANK repeat; Repeat.
SQ SEQUENCE 2087 AA; 218125 MW; 0955C33D7A7AA8F8 CRC64;

alignment_scores:
Quality: 127.50 Length: 124
Ratio: 1.903 Gaps: 9
Percent Similarity: 54.032 Percent Identity: 36.290

alignment_block:
US-09-462-480-2/rev x Q9WUE8 ..
Align seg 1/1 to: Q9WUE8 from: 1 to: 2087

353 CCCGGG.....CTGGTGGAGCGCGGAT...TGCAGACCCCTG 319
||||| ||| ::||| ||| ::|||
1512 ProGlyProHisProLeuProAspProSerProAlaThrProLe 1528
318 GCCATCGCTCCCGGACCC...ACCGAGCGCGCCACCCGTCACCGAGC 272
||| ||||| ||| ::||| ||| ||||| |||
1528 uProAlaProProProAlaValAlaAlaProProThrLeuAspS 1545
271 ATCCGGCAACAGCC.....GCCGGCATCACC 246
:: ||| ::|||
1545 erThrAlaSerSerLeuThrSerTyrAspSerGluValAlaThrLeuThr 1561
245 GAGGGGCAACCGGCTTTTCGATCAGCTGAGACATCAGCGCGTGGGGT 196
:: ||| |||
1562 GlnGlyAlaPro.....AlaAlaPr 1568
195 CAAGCACCCACCTGCGCCAGGT.....AGCGACTCCGCGCGCAGCGC 152
:: ||| ||| ||| |||
1568 oGlyAspProProAlaProGlyProProAlaProAlaProAlaProp 1585
151 CCGCGCGCGCGCTGGGCGCTGATCCACCGCAGCGGATGTTCCACAGC 102
||||| ||| ||||| ||| |||
1585 roAlaProGlnProGlyProAspProProGly.....ThrAspSer 1599
101 GGACTG.....GTCCCGCAGCAGG.....CCCATCTGCGC 73
||||| ||| |||||
1600 GlylleGluGluValAspSerSerSerSerSerHisProLeuGluTh 1616
72 GGCTTCCTCGTGGTGGTGGTGGT 51
: ||||| ::|||
1616 rIleSerSerAlaSerThrLeu 1623
```

seq_name: sp_rodent:Q9WU13

```

seq_documentation_block:
ID Q9WU13 PRELIMINARY; PRT; 2158 AA.
AC Q9WU13;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SYNAPTIC SAPAP-INTERACTING PROTEIN SYNAPSON.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99419021; PubMed=10488079;
RA Yao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;
RT "Synapson, a novel neuronal protein interacting with synapse-associated
RT protein 90/postsynaptic density-95-associated protein.";
RL J. Biol. Chem. 274:27463-27466(1999).
DR EMBL; AF102855; AAD04569.2; -.
DR HSSP; P00519; IABL.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR002965; P-rich_extensn.
DR InterPro; IPR001680; SAM.
DR InterPro; IPR001452; SH3.
```

```

353 CCGGG.....CTGGTGGAGCCGCGGAT....TGCGAACCCGTG 319
      ||| ::|||::: ::|
1583 ProGlyProHisProLeuproAspProSerProAlaThrProLeu 1599
      ||| ||||| ||| ::::::||| |||||::: |||:
318 GCCATCCTCCCGACCC...ACCAGGAGCGGCCACCGTCACCGAGC 272
      ||| ||||| ||| ::::::||| |||||::: |||:
1599 ProAlaIaIaProProAlaValAlaAlaIaIaProProThrLeuASP 1616
      ||| ||||| ||| ::::::||| |||||::: |||:
271 ATCCGGCAACAGCC.....CGCGCATCACGC 246
      :: |||::: |||
1616 erThraIaSerSerLeuThrSertyrAspSerGluValAlaThrLeuThr 1632
      ||| ||||| ||| ::::::||| |||||::: |||:
245 GAGGGGCAACGCGCTTTCGATCAGCTGAGACATCAGCGCGTCGGGTG 196
      ::||| |||||
1633 GlngLyAlaPro.....AAlaPr 1639
      ::||| |||||
195 CAAGACCCACTCGCGCAGGT.....AGCGACTCGCGCGCAGCAGGC 152
      ::||| ||||| ||| ::||| ::|
1639 oGlyAspProAlaProGlyProProAlaProAlaIaIaProAlaPro 1656
      ||| ||||| ||| ::::::||| |||||::: |||:
151 CGGCGCGCGCTGGGCTGTATCCACCACCGCGGTGTTCGACAGC 102
      ||||| ||||| ||||| ||||| ::|
1656 roAlaProGlInProGlyProAspProProGly.....ThrAspSer 1670
      ||| ||||| ||||| |||||
101 GGACTG.....GTGCGCAGCAGG.....CCCATCTGCCGC 73
      ||| |||||
1671 GlyleGluGluValaspSerArgSerSerSerAspHisproLeuLuth 1687
      ||| ||||| |||||
72 GCCTTCCTCGTCGCGTGGGTG 51
      ||| |||||::: |||
1687 rIleSerSerAlaSerThrLeu 1694

```

OM of: US-09-462-480-1 to: A_Geneseq_032802.* out_format : pfs
Date: Jul 22, 2002 1:22 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet.n2p.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09462480/runat_18072002.164418.19413/app_query.fasta_1.2850
-DB=A_Geneseq_032802-QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500
-FCGAP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480@cgn1.1_57 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-1
Query length: 1277
Database: A_Geneseq_032802.*
Database sequences: 747574
Database length: 111073796
Search time (sec): 134.850000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32452 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32384 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81705 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64338 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39135 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW38992 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW39844 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72929 +	713.00	1027.83	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW21946 +	713.00	1027.83	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW29888 +	512.50	741.37	1.9e-33		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW30705 +	492.00	718.28	7.3e-32		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW22942 +	487.50	697.48	2.6e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW21962 +	487.50	697.48	2.6e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81746 +	487.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64379 +	487.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW32063 +	487.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39224 +	487.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39176 +	487.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39081 +	487.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81706 +	486.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64339 +	486.00	709.51	2.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW39136 +	486.00	709.51	2.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW38993 +	486.00	709.51	2.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW35218 +	486.00	709.51	2.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW39845 +	486.00	709.51	2.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW1901 +	485.00	684.76	5.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72943 +	482.00	689.47	7.3e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW321963 +	482.00	689.47	7.3e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:AAW11494 +	479.00	699.81	8.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW29890 +	479.00	699.81	8.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW29788 +	479.00	699.81	8.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW35219 +	479.00	699.81	8.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32444 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32376 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81747 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64321 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW32097 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39118 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW38981 +	462.00	674.97	2.0e-29		

/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32454 + 392.00 574.44 9.4
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32386 + 392.00 574.44 9.4
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81707 + 392.00 574.44 9.4
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64340 + 392.00 574.44 9.4
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39137 + 392.00 574.44 9.4
seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32452

seq_documentation_block:

ID AAW32452 standard; Protein; 368 AA.
XX
AC AAW32452;
XX
DT 09-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen Tb37-FL.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 10 /note= "Any amino acid"
XX
PN WO9709428-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US14674.
XX
PR 12-JUL-1996; 96US-0680574.
PR 01-SEP-1995; 95US-0523436.
PR 22-SEP-1995; 95US-0533634.
PR 22-MAR-1996; 96US-0620874.
PR 05-JUN-1996; 96US-0659683.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skelky YAW;
PI Twardzik DR, Vedvick TH;
XX
DR WPI; 1997-192903/17.
XX
PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
PT useful in vaccines for prevention or treatment of tuberculosis, also
PT for diagnosis

Example 3; Page 146-147; 168pp; English.

A new immunogenic polypeptide has been developed comprising an immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its variant differing only in conservative substitutions and/or modifications). The present sequence represents a M.tuberculosis antigen, Tb37-FL The immunogenic protein, and fusion proteins containing one or more of the proteins or one of the proteins plus ESAT-6, are useful in vaccines, preferably when formulated with a non-specific adjuvant, to induce an immune response against M.tuberculosis (for treatment or prevention).

Sequence 368 AA;

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
alignment_block:
US-09-462-480-1 x AAW32452 ..
Align seg 1/1 to: AAW32452 from: 1 to: 368

```
1  CTGACGAGGTGACGTCGTGTTTCACCCAGGTGGCGCGCACCGCGCGG 50
|||||
226  LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
|||||
51  CAACCCAGCGGAGGAGGAGCGGAGATGGCCCTGCTCGGCACCAAGTC 100
|||||
242  yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101  CGCTGTGCAACCATCGCTGGTGGTGGATCAGGCCAGCCAGCGCGCGG 150
|||||
259  roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151  GCGTCTGCTCGCGGAGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 200
|||||
276  GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201  CACGCGCTGATCTCTCAGCTGATCGAAAGCGGTTGCCCTCGGTGA 250
|||||
292  gThrProLeuMetSerGlnLeuLeuGlnProValAlaProSerValM 309
|||||
251  TCGCGCGGCTGTCGCGGATCGTGGTGGTGGTGGTGGTGGTGGTGG 300
|||||
309  etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaPro 325
|||||
301  GGTCCGGGAGCGATGGCGGAGGTTCGCAATCCGGCGGCTCCACGACC 350
|||||
326  GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
|||||
351  GGTCTGCTGCGCGCGGCGGCGGCTGCGCGGAGCGGTGAGAGACGAG 400
|||||
342  GglyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
|||||
401  AGGAGACTGGGACGAAGAGGACGACTGG 429
|||||
359  luAspAspTrpAspGluGluAspAspTrp 368
|||||
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32384

seq_documentation_block:

ID_AAW32384 standard; Protein; 368 AA.

AC_AAW32384;

13-JAN-1998 (first entry)

Mycobacterium tuberculosis antigen Tb37-FL.

Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
skin testing; M.tuberculosis.

Mycobacterium tuberculosis.

Key Location/Qualifiers

Misc_difference 10 /note= "Any amino acid"

WO9709429-A2.

13-MAR-1997.

30-AUG-1996; 96WO-US14675.

12-JUL-1996; 96US-0680573.

01-SEP-1995; 95US-0523435.

22-SEP-1995; 95US-0532136.

22-MAR-1996; 96US-0620280.

03-JUN-1996; 96US-0658800.

(CORI-) CORIXA CORP.

Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI_Twardzik DR, Vedvick TH;

```
XX WPI; 1997-192904/17.
XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
XX - useful for diagnosis of M. tuberculosis infection
XX Example 3; Page 159-161; 190pp; English.
XX A new immunogenic polypeptide has been developed comprising an
XX immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
XX its variant differing only in conservative substitutions and/or
XX modifications). The present sequence represents a M.tuberculosis
XX antigen, Tb37-FL. The immunogenic polypeptide can be used to diagnose
XX M.tuberculosis infection by forming complexes with specific
XX antibodies in the sample. Fragments of DNA encoding the immunogenic
XX polypeptide can be used as diagnostic primers or probes and agents
XX that bind to the antigen, especially monoclonal antibodies or
XX equivalent polyclonal antibodies, are also used for diagnosis.
XX Sequence 368 AA;
```

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-1 x AAW32384

Align seg 1/1 to: AAW32384 from: 1 to: 368

```
1  CTGACGAGGTGACGTCGTGTTTCACCCAGGTGGCGCGCACCGCGCGG 50
|||||
226  LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
|||||
51  CAACCCAGCGGAGGAGGAGCGGAGATGGCCCTGCTCGGCACCAAGTC 100
|||||
242  yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101  CGCTGTGCAACCATCGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 150
|||||
259  roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151  GCGTCTGCTCGCGGAGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 200
|||||
276  GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201  CACGCGCTGATCTCTCAGCTGATCGAAAGCGGTTGCCCTCGGTGA 250
|||||
292  gThrProLeuMetSerGlnLeuLeuGlnProValAlaProSerValM 309
|||||
251  TCGCGCGGCTGTCGCGGATCGTGGTGGTGGTGGTGGTGGTGGTGG 300
|||||
309  etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaPro 325
|||||
301  GGTCCGGGAGCGATGGCGGAGGTTCGCAATCCGGCGGCTCCACGACC 350
|||||
326  GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
|||||
351  GGTCTGCTGCGCGGCGGCGGCTGCGCGGAGCGGTGAGAGACGAG 400
|||||
342  GglyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
|||||
401  AGGAGACTGGGACGAAGAGGACGACTGG 429
|||||
359  luAspAspTrpAspGluGluAspAspTrp 368
|||||
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81705

seq_documentation_block:

ID_AAW81705 standard; Protein; 368 AA.

XX


```

AC AAW81705;
XX
XX 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide Tb37-FL.
XX
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
XX Mycobacterium tuberculosis.
XX
XX Key Location/Qualifiers
FT Misc-difference 10
FT /label= unknown
XX
XX WO9816646-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18293.
XX
XX 13-MAR-1997; 97US-0818112.
XX 11-OCT-1996; 96US-0730510.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX WPI; 1998-261042/23.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
XX to develop products for the detection of M. tuberculosis infection
XX and for diagnosis, treatment and prevention of tuberculosis
XX
XX Example 3B; Page 137-138; 230pp; English.
XX
XX This sequence represents an immunogenic portion of a soluble
XX Mycobacterium tuberculosis (MT) antigen which can be used in a method
XX for inducing protective immunity against tuberculosis (TB). This
XX sequence can be formulated into vaccines and/or pharmaceutical
XX compositions for immunising against M. tuberculosis infection or may
XX be used for the diagnosis of tuberculosis.
XX
XX Sequence 368 AA;

alignment_scores:
    Quality: 713.00      Length: 143
    Ratio: 5.057         Gaps: 0
    Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-1 x AAW81705 ..

Align seq 1/1 to: AAW81705 from: 1 to: 368

1 CTGCAGCAGGTGAGCTGTTGTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
|||||
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCGCTCGGCACCAAGTC 100
|||||
242 YASNProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGACACCATCCGCTGGTGGATGATGAGCCCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGCTGGCGCGGAGTGGCTACCTGGCGCAGGTGGTGGTGGACCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292

```

```

201 CAGCGCGCTGATGTCTCAGCTGATCGAAAGCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGCCGGCGGCTGTTCCGGATCGTTCGGTGACGGTGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGGAGCGATGGCCAGGTTCCGAATCGCGGCTCCACCGACCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGCTGCTGGTCGCGCGCACCGCTCGCGCAGGAGCGGTGAAGAAGACGAC 400
|||||
342 OGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGAGGAGAGGAGGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64338
seq_documentation_block:
ID AAW64338 standard; Protein: 368 AA.
XX
XX AAW64338;
XX
XX 09-NOV-1998 (first entry)
XX
XX Mycobacterium tuberculosis antigen Tb37-FL.
XX
XX Tuberculosis; infection; diagnosis; antigen; Tb37-FL.
XX
XX Mycobacterium tuberculosis strain H37Rv.
XX
XX Key Location/Qualifiers
FT Misc-difference 10
FT /note= "unidentified"
XX
XX WO9816645-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18214.
XX
XX 13-MAR-1997; 97US-0818111.
XX 11-OCT-1996; 96US-0729622.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX WPI; 1998-251292/22.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX to develop products for the detection of M. tuberculosis infection
XX and diagnosis of tuberculosis
XX
XX Example 3; Page 143-144; 250pp; English.
XX
XX This polypeptide comprises Mycobacterium tuberculosis antigen
XX Tb37-FL. It is encoded by genomic DNA isolated from a M.
XX tuberculosis strain H37Rv genomic library using a probe from
XX clone Tb38-1 (see AAV44384). The invention relates to compositions
XX and methods for diagnosing tuberculosis. It provides polypeptides
XX (see AAW64291-W64379) comprising an antigenic portion of a soluble
XX M. tuberculosis antigen, or an immunogenic portion of an M.
XX tuberculosis antigen, as well as DNA sequences encoding such
XX polypeptides, recombinant expression vectors and transformed or
XX transfected host cells. Also claimed are methods and diagnostic
XX kits for detecting M. tuberculosis infection in a patient using
XX these polypeptides, antibodies or oligonucleotide probes and

```

CC primers, for the diagnosis of tuberculosis.

XX
SQ Sequence 368 AA;

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-1 x AAW64338 ..

Align seg 1/1 to: AAW64338 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTCGCTGTTTCAGCCAGGTGGCGGCACCGCGGGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG1 242
51 CAACCCAGCCGACGAGGAGCGGCGCAGATGGGCTGCTCGGCACCAATC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCCGTCGCTGGTGTGATCAGCCCGCCAGCGGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGGGAGTCTGCTACCTGGCGCAGGTGGTGTGTCACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATGTCATCAGTGTGATCGAAAGCCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTGCGGATCGCTGGTGACGGGTGGCGCGCTCGCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCGGGAGGATGGCGGCGGTTCGCAATCCGCGCGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGGTCGCGCGGCGACCGCTCGCGCAGGACCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGACGACTGGGACGAGGACGACTGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.AAY39135

seq_documentation_block:

ID AAY39135 standard; Protein; 368 AA.

XX
AC AAY39135;

XX
DT 05-NOV-1999 (first entry)

XX
DE M. tuberculosis antigen Tb37-FL amino acid sequence.

XX
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.

XX
OS Mycobacterium tuberculosis.

XX
PN WO9942076-A2.

XX
PD 26-AUG-1999.

XX
PF 17-FEB-1999; 99WO-US03268.

XX
PR 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;
XX
DR WPI; 1999-527409/44.
XX
PT New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX
PS Example 3; Page 132-133; 299pp; English.
XX
CC The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA19249 to AA19460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.
XX
SQ Sequence 368 AA;

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-1 x AAY39135 ..

Align seg 1/1 to: AAY39135 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTCGCTGTTTCAGCCAGGTGGCGGCACCGCGGGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG1 242
51 CAACCCAGCCGACGAGGAGCGGCGCAGATGGGCTGCTCGGCACCAATC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCCGCTGGTGTGATCAGCCCGCCAGCGCGGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGGCGGAGTCTGCTACCTGGCGCAGGTGGTGTGTCACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATGTCATCAGTGTGATCGAAAGCCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTGCGGATCGCTGGTGTGATCAGCGCGCTCCCGCTCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCGGGAGGATGGCGGCGGTTCGCAATCCGCGCGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGGTCGCGCGGCGACCGCTCGCGCAGGACCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
```

259 r0LeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

CC prophylactic (vaccine) methods, including genetic vaccination methods.

```
XX SQ Sequence 368 AA;
alignment_scores:
  Quality: 713.00 Length: 143
  Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
alignment_block:
US-09-462-480-1 x AAB19844 ..
Align seg 1/1 to: AAB19844 from: 1 to: 368
1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCGCGGGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG1 242
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGCTGCTCGGCACCAATC 100
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCAACCATCCGCTGCTGCTGATCAGCCAGCCAGCGCGGCG 150
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTCTGCGCGGAGTCTGCTACCTGGCGCAGGTGGTGGTTCAC 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
201 CACGCGCTGATGCTCTAGCTGATCGAAGCCGGTTCGCCCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309
251 TGCCGCGGCTGTTCGCGGATCTGCTGCTGACGGTGGCGGCTCCGGT 300
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGGAGCGATGGCCAGGTTTCGCAATCCGCGGCTCCACAGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrAr 342
351 GGGTCTGGTCTGCGCGGCGGCGGCTCGCGCAGGAGCGGTGAAGACG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
401 AGGACGACTGGGACGAGAGGAGGACTGG 429
359 LuAspAspTrpAspGluGluAspTrp 368
seq_name: /STDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72929
seq_documentation_block:
ID AAW72929 standard; Protein; 371 AA.
XX AC AAW72929;
XX DT 21-JAN-1999 (first entry)
XX DE
XX DE Mycobacterium tuberculosis antigen RD1-ORF5.
XX KW Mycobacterium tuberculosis; antigen; vaccine; immunological;
XX KW immunogen; infection.
XX OS
XX OS Mycobacterium tuberculosis.
XX PN
XX PN W09844.1119-Al.
XX PD
XX PD 08-OCT-1998.
XX PF
XX PF 01-APR-1998; 98WO-DK00132.
XX PX
XX PX 05-JAN-1998; 98US-0070488.
PR 02-APR-1997; 97DK-0000376.
PR 18-APR-1997; 97US-0044624.
PR 10-NOV-1997; 97DK-0001277.
XX (STAT-) STATENS SERUM INST.
XX Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB;
PI Rosenkrands I, Weidinger K;
XX WPI; 1998-542705/46.
DR N-PSDB; AAV63939.
XX
PT New isolated mycobacteria polypeptides and nucleic acids - used for
PT developing products for the diagnosis of or vaccination against
PT mycobacterial infections, particularly tuberculosis
XX Claim 1; Page 200-202; 163pp; English.
XX
CC The present sequence represents a Mycobacterium tuberculosis protein.
CC Products from the present invention, which describes protein fragments
CC and nucleic acid fragments derived from M.tuberculosis, can be used in
CC the detection of and prevention of mycobacterial infections. In
CC particular, the proteins and nucleic acids can be used for the diagnosis
CC of or vaccination against tuberculosis caused by M. tuberculosis,
CC M. africanum or M. bovis.
XX
SQ Sequence 371 AA;
alignment_scores:
  Quality: 713.00 Length: 143
  Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
alignment_block:
US-09-462-480-1 x AAW72929 ..
Align seg 1/1 to: AAW72929 from: 1 to: 371
1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCGCGGGG 50
229 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG1 245
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGCTGCTCGGCACCAATC 100
245 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 262
101 CGCTGTCAACCATCCGCTGCTGCTGATCAGCCAGCCAGCGCGGCG 150
262 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 278
151 GGCCTGTCTGCGCGGAGTCTGCTACCTGGCGCAGGTGGTGGTTCAC 200
279 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 295
201 CACGCGCTGATGCTCTAGCTGATCGAAGCCGGTTCGCCCTCGGTGA 250
295 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 312
251 TGCCGCGGCTGTTCGCGGATCTGCTGCTGATCAGCGGTGGCGGCT 300
312 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 328
301 GGTCCGGGAGCGATGGCCAGGTTTCGCAATCCGCGGCTCCACAGCC 350
329 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrAr 345
351 GGGTCTGGTCTGCGCGGCGGCGGCTCGCGCAGGAGCGGTGAAGACG 400
345 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 362
401 AGGACGACTGGGACGAGAGGAGGACTGG 429
```

362 luAspAspTrpAspGluAspAspTrp 371

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV21946

seq_documentation_block:

ID AAY21946 standard; Protein; 371 AA.

AC AAY21946;

DT 06-SEP-1999 (first entry)

DE Amino acid sequence of antigen RD1-ORF5.

XX Immunogenic; Mycobacterium tuberculosis; immune response; infection;
KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;
KW pharmaceutical; vaccination; M. africanus; M. bovis; CFP7A; CFP30A;
KW CFP7B; CFP19; CFP27; CFP30A; RD1-ORF; CFP10A; CFP16; CFP19; CFP23;
KW CFP25A; CFP30B; CFP7B.

XX Mycobacterium tuberculosis.

OS WO9924577-A1.

PN 20-MAY-1999.

PD 08-OCT-1998; 98WO-DK00438.

PF 01-APR-1998; 98WO-DK00132.

PR 10-NOV-1997; 97DK-0001277.

PR 05-JAN-1998; 98US-0070488.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Skjot R;

XX WPI; 1999-347282/29.

XX N-PSDB; AAX81046.

XX New immunogenic fragment of Mycobacterium tuberculosis

XX Example 2; Page 219-220; 265pp; English.

XX The invention describes a substantially pure immunogenic polypeptide
CC fragment (1) from Mycobacterium tuberculosis that is able to evoke a
CC protective immune response against infections by mycobacteria belonging
CC to the tuberculosis complex. The invention provides a (1) fusion
CC polypeptide comprising at least one polypeptide fragment (1) and at least
CC one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell
CC epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second
CC different amino acid sequence from M. tuberculosis, and/or including a
CC sequence which protects the first amino acid sequence from in vivo
CC degradation or post-translational processing; (3) a nucleic acid fragment
CC that encodes the above polypeptides. The polypeptides and nucleic acid
CC are useful as pharmaceuticals, for diagnosis of and as antigens for
CC vaccination against TB caused by Mycobacterium tuberculosis, africanus or
CC bovis. The polypeptides are also useful for diagnosing ongoing or
CC previous sensitization in an animal with bacteria belonging to the
CC tuberculosis complex. The invention also describes the use of CFP7A or
CC CFP30A or a T-cell epitope of for the induction of a strong immune
CC response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell
CC epitope of for diagnosis of TB in a mammal by performing a DTH type skin
CC test; use of CFP27, CFP30A, RD1-ORF2, RD1-ORF3, RD1-ORF5, MPT59-ESAT6,
CC ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-
CC cell epitope of for the preparation of an immunological composition; and
CC for the preparation of a subunit vaccine.

XX Sequence 371 AA;

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-1 x AAY21946 ..

Align seg 1/1 to: AAY21946 from: 1 to: 371

1 CTGACGAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCCGCGCGCG 50
|||||
229 LeuGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 245
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGGCTGCTGGCACCAGTC 100
|||||
245 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 262
101 CGCTGTCAACCATCCGCTGGTGGTATCAGCCAGCCCGCGCGCGCG 150
|||||
262 rOleuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 278
151 GGCCTGTGCGCGCGGAGTCTACCTGCGCAGGTGGTGGTCTGTGACCCG 200
|||||
279 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 295
201 CACGCCGCTGATGTCCTCAGCTGATCGAAAGCCGTTGCCCTCGGTGA 250
|||||
295 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 312
251 TGCCGCGCGCTGTTTGGCGGATCGTGGTACGGTGGCGCGCTCCGGTG 300
|||||
312 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 328
301 GGTCGCGGAGCGATGGCCAGGTTCGCAATCCGCGCGCTCCACCCAGCC 350
|||||
329 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 345
351 GGGTCTGTCGCGCGCGCGCGCTCGCGCAGGCGCTGAGAAAGACGACG 400
|||||
345 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 362
401 AGGACGACTGGGACGAGGAGGACGACTGG 429
362 luAspAspTrpAspGluAspAspTrp 371

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV29888

seq_documentation_block:

ID AAY29888 standard; Protein; 196 AA.

XX AAY29888;

XX 18-NOV-1999 (first entry)

XX Mycobacterium tuberculosis ESAT-6 protein sequence.

XX Mycobacterial; lactic acid bacterium; diagnosis; skin test; vaccine;
KW delayed type hypersensitivity; DTH; ESAT-6 homodimer; tuberculosis;
KW interferon-gamma release.

XX Mycobacterium tuberculosis.

XX WO9945119-A2.

XX 10-SEP-1999.

XX 05-MAR-1999; 99WO-DK00109.

XX 06-MAR-1998; 98DK-0000306.

XX 06-MAR-1998; 98US-0077105.

XX (STAT-) STATENS SERUM INST.

XX Jensen CL, Folkersen J;

XX WPI; 1999-551043/46.

211 gAsnAspProThrGlnGlnMileProLysLeuValAlaAsnAsnThrArgL 228
569 AGGTAATTCGAGCGGATCT 588
228 autrpValTyrCysGlyAsnGlyThrProAsnGluLeuGlyGlyAlaAsn 244
589 ...CCGGGACCTGAAACCCAGATCGACCGAGGTGCGACGCGAGGT 635
245 ileProAlaGluPheLeuGluAsnGlyValArgSerSerAsnLeuLysph 261
636 TCGTTTCAGGGCCAGTGGCGCGCGCGCGGACGG C 673
261 eGlnAspAlaTyrAsnAlaAlaGlyGlyHisAsnAlaValPheAsnPhP 278
674 CGCCGAGCGCGGTGGTGGCTTCCAAAGACGACGCAATAAGCAGAAGC 723
278 toPro. 279
724 AGGAACCTCGACGAGATCTCGACGAATATTCTGTCAGCGCGCGTCCAATAC 773
280 ...AsnGlyThrHisSerTyrGluTyrTrpGlyAlaGlnLeuAsnAlaM 295
774 TCGAGGCGCGGAGGAGCAGCAGCGCTGCTCTCGCAATGGGCTT 823
295 eLysGlyAspLeuGlnSerLeuGlyAla GlyLys 307
824 CTGACCGCGCTAATACGAAAGAAAGACGAGCAAAACATGACAGACGCA 873
308 LeuAla MetThrGluGlnG 314
874 GTGGAAATTCGGGGTATCGAGCGCGCGCAAGCGCAATCCAGGAAATG 923
314 nTrpAsnPheAlaGlyIleGluAlaAlaAlaSerAlaIleGlnGlyAsnV 331
924 TCACCTCCATTCCTCTCTGACGAGGGAAGCAGTCCCTGACCAAG 973
331 alrSerIleHisSerLeuAspGlyGlyLysGlnSerLeuThrLys 347
974 CTCGAGCGCGCTGGGGGAGTTCGAGGCGGTACCGAGGTGTGCA 1023
348 LeuAlaAlaAlaTrpGlySerGlySerGluAlaTyrGlnGlyValG 364
1024 GCAAAATGGAGCGCCAGCGGTACCGAGCTGAACACGCGTGCAGAAC 1073
364 nGlnLysTrpAspAlaThrAlaThrGluLeuAsnAsnAlaLeuGlnAsnL 381
1074 TGGCGCGGACGATCAGCAGCGGTACGCAATGCTTCGACCGCAAGGC 1123
381 euAlaArgThrIleSerGluAlaGlyGlnAlaMetAlaSerThrGluGly 397
1124 AACGTCACCTGGGATGTCGCA 1144
398 AsnValThrGlyMetPheAla 404

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AA1999

seq_documentation_block:

ID_AAY21962 standard; Protein; 404 AA.

XX AC AAY21962;

XX DT 06-SEP-1999 (first entry)

XX DE Amino acid sequence of antigen MPT59.

XX KW Immunogenic; Mycobacterium tuberculosis; immune response; infection;
KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;
KW pharmaceutical; vaccination; M. africanum; M. bovis; CFP30A;
KW CFP7B; CFP19; CFP27; CFP30A; RDI-ORF; CFP16; CFP19; CFP23;
KW CFP25A; CFP30B; CFP7B.

XX OS Mycobacterium tuberculosis.

XX

PN W09924577-A1.
XX 20-MAY-1999.
XX 08-OCT-1998; 98WO-DK00438.
XX 01-APR-1998; 98WO-DK00132.
PR 10-NOV-1997; 97DK-0001277.
XX 05-JAN-1998; 98US-0070488.
XX (STAT-) STATENS SERUM INST.
XX Andersen P, Skjot R;
XX WPI; 1999-347282/29.
XX New immunogenic fragment of Mycobacterium tuberculosis
PT
XX
XX Examples; Page 249-250; 265pp; English.

The invention describes a substantially pure immunogenic polypeptide fragment (I) from Mycobacterium tuberculosis that is able to evoke a protective immune response against infections by mycobacteria belonging to the tuberculosis complex. The invention provides a (1) fusion polypeptide comprising at least one polypeptide fragment (I) and at least one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second different amino acid sequence from M. tuberculosis, and/or including a sequence which protects the first amino acid sequence from in vivo degradation or post-translational processing; (3) a nucleic acid fragment that encodes the above polypeptides. The polypeptides and nucleic acid are useful as pharmaceuticals, for diagnosis of and as antigens for vaccination against TB caused by Mycobacterium tuberculosis, africanum or bovis. The polypeptides are also useful for diagnosing ongoing or previous sensitization in an animal with bacteria belonging to the tuberculosis complex. The invention also describes the use of CFP7A or CFP30A or a T-cell epitope of for the induction of a strong immune response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell epitope of for diagnosis of TB in a mammal by performing a DTH type skin test; use of CFP27, CFP30A, RDI-ORF2, RDI-ORF3, MPT59-ESAT6, ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-cell epitope of for the preparation of an immunological composition; and for the preparation of a subunit vaccine.

XX Sequence 404 AA;

alignment_scores:

Quality: 487.50 Length: 391

Ratio: 2.579 Gaps: 13

Percent Similarity: 48.338 Percent Identity: 36.829

alignment_block:

US-09-462-480-1 x AAY21962 ..

Align seg 1/1 to: AAY21962 from: 1 to: 404

328 CAATCCGGCGGCTCCACCAGCGCGGTCTGGTCGCGCGCGACCGCTCGC 377
|||||

48 GlnSerGlyGlyAsnAsnSerProAlaValTyrLeu LeuAs 61

378 GCAGGAGCGTGAAGAAGACGAGCGAGCTGGGACGAAGAG 420
|||

61 pGlyLeuArgAlaGlnAspAspTyrAsnGlyTrpAspIleAsnThrProA 78

420 420

78 laPheGluTrpTyrGlnSerGlyLeuSerIleValMetProValGly 94

421GAGACTGGTGAGCTCCCGTA 441
:::|||||

95 GlyGlnSerSerPheTyrSerAspTrpTyrSerProAlaCysGlyLysAl 111

442ATGACACAGACTTCCCG. 459
111 aGlyCysGlnThrTyrlsYrtpGluThrPheLeuThrSerGluLeuProG 128
459 459
128 lnTrpLeuSerAlaAsnArgAlaVallysProThrGlySerAlaAlaile 144
460 ...GCCACCGCGCGCAAGACTTCCCAACATTTTGGCGGAGGAGTAA 506
145 GlyLeuSerMetAlaGlySerSerAlaMetIleLeuAlaIaTyrHisPr 161
507 GAGAGAA..... 513
161 oGlnGlnPheIleTyrrAlaGlySerLeuSerAlaLeuLeuAspProSerG 178
513 513
178 lnGlyMetGlyProSerLeuIleGlyLeuAlaMetGlyAspAlaGlyGly 194
514AGTAGTCCAGCATGGCAGAG 533
195 TyrlsAlaAlaaspMetTrpGlyProSerSerAspProIaIrrpGluAr 211
534 ATGAGAGCG.....ATGCGG.....CTACCCCTCGGCGAGGAGGC 568
211 gAsnAspProThrGlnGlnIleProLysLeuValAlaAsnThrArgL 228
569 AGTAATTCGAGCGGATCT..... 588
228 euTrpValTyrcysGlyAsnGlyThrProAsnGluLeuGlyGlyAlaAsn 244
589 ...CCGCGCAGCTGAAACCCAGATCCAGCAGGTGGAGTCGACGCGAGGT 635
245 IleProAlaGluPheLeuGluAsnPheValArgSerSerAsnLeuLysPh 261
636 TCGTTCGAGCGCCAGTCGCGCGCGCGCGGACGG.....C 673
261 eGlnAspAlaTyrrAsnAlaAlaGlyGlyHisAsnAlaValPheAsnPh 278
674 CGCCCGCGCGGTGTGTGCTTCCAGAGACGACCAATAAGCAGAGC 723
278 roPro..... 279
724 AGGAACCTGACGAGATCTCGACGAATATTCGTACGCGCGGTCCAATAC 773
280 ...AsnGlyThrHisSerTrp.GluTyrrTrpGlyAlaGlnLeuAsnAlaM 295
774 TCGAGGCGCGCAGGAGCAGCAGCAGCGCTGTCTCGCAATPGGCTT 823
295 etLysGlyAspLeuGlnSerSerLeuGlyAla.....GlyLys 307
824 CTGACCGCTTAATACGAAGAAACGAGCAACATGACAGAGCAGCA 873
308 LeuAla.....MetThrGluGlnG 314
874 GTGAATTTTCGGGTATCGAGCGCGCAAGCGCAATCCAGGAAATG 923
314 nTrpAsnPheAlaGlyIleGluAlaAlaAlaSerAlaIleGlnGlyAsnV 331
924 TCAGCTCATTCATTCCTCTCTGACGAGGGAAGCAGTCCTCGACCAAG 973
331 alThrSerIleHisSerLeuLeuAspGluGlyLysGlnSerLeuThrLys 347
974 CTCGCGCGCGCTGGCGGTAGCGGTTCGAGCGGTACCGAGGTGTCCA 1023
348 LeuAlaAlaIaIaTrpGlyGlySerGlySerGluAlaTyrrGlnGlyValG 364
1024 GCAAAATGGAGCGCGCTACCGAGCTGAACAACGCGCTCGAGAAC 1073
364 nGlnLysTrpAspAlaThrAlaThrGluLeuAsnAlaLeuGlnAsnL 381
1074 TGGCGCGCAGCATCAGCGAAGCGCGTCAGGCAATGGCTTCGACCGAAGGC 1123

|||||
381 euAlaArgThrIleSerGluAlaGlyGlnAlaMetAlaSerThrGluGly 397
1124 AACGTCACTGGGATGTCGCA 1144
398 AsnValThrGlyMetPheAla 404
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.AAW81746
seq_documentation_block:
ID AAW81746 standard; Protein; 802 AA.
XX AAW81746;
XX AC
XX 27-JAN-1999 (first entry)
XX DE M. tuberculosis fusion protein TbF-2.
XX KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX KW vaccine; pharmaceutical; infection; diagnosis.
XX OS Synthetic.
XX OS Mycobacterium tuberculosis.
XX PN W09816646-A2.
XX PD 23-APR-1998.
XX PF 07-OCT-1997; 97WO-US18293.
XX PR 13-MAR-1997; 97US-0818112.
XX PR 11-OCT-1996; 96US-0730510.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI: 1998-261042/23.
XX DR N-PSDB; AAV64567.
XX PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
XX PS Disclosure; Page 208-211; 230pp; English.
XX CC This sequence represents the fusion protein TbF-2 which is composed of
CC immunogenic polypeptides from Mycobacterium tuberculosis (MT). This
CC protein is used in a method for inducing protective immunity against
CC tuberculosis (TB). This sequence can be formulated into vaccines
CC and/or pharmaceutical compositions for immunising against
CC M. tuberculosis infection or may be used for the diagnosis of TB.
XX SQ Sequence 802 AA;

alignment_scores:
Quality: 487.00 Length: 103
Ratio: 4.822 Gaps: 0
Percent Similarity: 98.058 Percent Identity: 97.087
alignment_block:
US-09-462-480-1 x AAW81746
Align seg 1/1 to: AAW81746 from: 1 to: 802
522 AGCATGCGAGATGAAGACCGATCCCGCTACCTCGGCGAGGAGCAGG 571
423 SerSerAlaGluMetLysThrAspAlaThrLeuAlaGlnGluAlaG 439
572 TAATTTTCGAGCGGATCTCCGCGCGACCTGAAACCCAGATCGACCGGTGG 621
|||||

439 yAsnPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValG 456
622 AGTCAGCGCAGGTGCTGTCAGGCGCCAGTGGCGCGCGCGGGGACG 671
|||||
456 luSerThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 472
672 GCCGCCAGCGCGGGTGGTGGCTTCCAAAGCAGCAGCCAAATAGCAGAA 721
|||||
473 AlaAlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLy 489
722 GCAGGAACCTGCAGCAGATCTCGAGCAATATTCGTACGCGCGCGTCCCAAT 771
|||||
489 sGlnGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnT 506
772 ACTCGAGCGCGCAGCAGCAGCAGCGCGCTGTCTCGCAAAATGGCG 821
|||||
506 yrSerArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGly 522
822 TTCTGACCC 830
|||||
523 PheValPro 525

seq_name: /SDSI/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT.AAW64379.

seq_documentation_block:

ID AAW64379 standard; Protein; 802 AA.

XX AAW64379;

XX AC

XX DT 09-NOV-1998 (first entry)

XX DE Mycobacterium antigen Tbf2 protein fusion.

XX KW Tuberculosis; infection; diagnosis; 38 kDa antigen; Tbra3; DPEP;

XX KW Tbf38-1; Tbf-2.

XX OS Mycobacterium tuberculosis.

XX OS Synthetic.

XX PN WO9816645-A2.

XX PD 23-APR-1998.

XX PF 07-OCT-1997; 97WO-US18214.

XX PR 13-MAR-1997; 97US-0818111.

XX PR 11-OCT-1996; 96US-0729622.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX DR WPI; 1998-251292/22.

XX DR N-PSDB; AAV55801.

XX PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used

XX PT to develop products for the detection of M. tuberculosis infection

XX PT and diagnosis of tuberculosis

XX PS Example 7; Page 223-226; 250pp; English.

XX CC This polypeptide comprises a fusion protein, designated Tbf-2,
CC composed of Mycobacterium tuberculosis antigens Tbra3 (see AAW64295),
CC 38 kDa antigen (see AAW64364), Tbf38-1 (see AAW64321) and DPEP (see
CC AAW64322). It was produced by PCR amplification (see AAV44450-57) of
CC the appropriate antigen DNA sequences, cloning into an expression
CC vector, and expression in E. coli. Tbf-2 can be used for
CC serodiagnosis of tuberculosis. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAW64291-W64379) comprising antigenic or
CC immunogenic portions of M. tuberculosis antigens, or fusion proteins,
CC DNA sequences encoding such polypeptides, recombinant expression

CC vectors and host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient.

XX

SQ Sequence 802 AA;

alignment_scores:

Quality: 487.00 Length: 103

Ratio: 4.822 Gaps: 0

Percent Similarity: 98.058 Percent Identity: 97.087

alignment_block:

US-09-462-480-1 x AAW64379

Align seg 1/1 to: AAW64379 from: 1 to: 802

522 AGCATGCGCAGAGATCAAGACCGATGCCCTACCTCGGGCAGGAGGAGG 571
|||||
423 SerSerAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGln 439
572 TAATTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCAAGTGG 621
|||||
439 yAsnPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValG 456
622 AGTCAGCGCAGGTGCTGTCAGGCGCCAGTGGCGCGCGCGGGGACG 671
|||||
456 luSerThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 472
672 GCCGCCAGCGCGGGTGGTGGCTTCCAAAGCAGCAGCCAAATAGCAGAA 721
|||||
473 AlaAlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLy 489
722 GCAGGAACCTGCAGCAGATCTCGAGCAATATTCGTACGCGCGCGTCCCAAT 771
|||||
489 sGlnGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnT 506
772 ACTCGAGCGCGCAGCAGCAGCAGCAGCGCTGTCTCGCAAAATGGCG 821
|||||
506 yrSerArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGly 522
822 TTCTGACCC 830
|||||
523 PheValPro 525

OM of: US-09-462-480-1 to: Issued_Patents_AA:* out_format : pfs
Date: Jul 22, 2002 1:24 AM
About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O/cgn2_1/USPTO.spool/US09462480/runat_18072002_164418_19425/app_query.fasta_1.2850
-DB=Issued_Patents_AA -OPMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.500
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blousum62
-TRANS-human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480 -CGN1_18 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT_THREADS=1

Search information block:

Query: US-09-462-480-1
Query length: 1277
Database: Issued_Patents_AA:*
Database sequences: 231628
Database length: 24425594
Search time (sec): 53.850000

score_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-112-114	+	713.00	1026.33	3.0e-50	368	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-111-109	+	713.00	1026.33	3.0e-50	368	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-056-556-114	+	713.00	1026.33	3.0e-50	368	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-056-556-114	+	487.00	687.90	9.7e-32	802	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-112-115	+	486.00	708.68	5.4e-32	100	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-111-110	+	486.00	708.68	5.4e-32	100	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-056-556-115	+	486.00	708.68	5.4e-32	100	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-465-640-2	+	479.00	699.01	2.0e-31	95	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-112-88	+	462.00	674.18	4.8e-30	95	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-111-89	+	462.00	674.18	4.8e-30	95	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-056-556-88	+	462.00	674.18	4.8e-30	95	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-112-117	+	392.00	573.77	2.2e-24	80	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-111-112	+	392.00	573.77	2.2e-24	80	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-056-556-117	+	392.00	573.77	2.2e-24	80	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-112-104	+	259.00	384.32	1.2e-13	51	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-111-99	+	259.00	384.32	1.2e-13	51	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-056-556-104	+	259.00	384.32	1.2e-13	51	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-199-637A-277	+	209.50	286.46	3.2e-09	558	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-413-814-78	+	181.00	239.94	7.9e-07	882	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-07-945-283-2	+	180.00	238.48	9.6e-07	882	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-041-886-23	+	165.00	213.42	1.8e-05	1185	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-413-814-79	+	164.00	211.71	2.2e-05	1213	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-483-533-41	+	160.50	219.72	2.6e-05	355	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-283-471A-41	+	160.50	219.72	2.6e-05	355	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-063-533-41	+	160.50	219.72	2.6e-05	355	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-07-928-611-13	-	159.00	220.57	3.1e-05	267	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-487-811A-13	-	159.00	220.57	3.1e-05	267	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-060-694-13	-	159.00	220.57	3.1e-05	267	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-060-694-13	-	159.00	220.57	3.1e-05	267	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-056-051-6	-	159.00	214.60	3.9e-05	467	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-07-928-611-22	-	159.00	214.60	3.9e-05	467	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-487-811A-22	-	159.00	214.60	3.9e-05	467	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-060-694-22	-	159.00	214.60	3.9e-05	467	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-060-694-22	-	159.00	214.60	3.9e-05	467	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-060-694-22	-	159.00	214.60	3.9e-05	467	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-931-820-2	+	156.00	201.83	9.1e-05	1024	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-963-825-19	+	156.00	198.75	0.0001	1366	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-570-573-19	+	156.00	198.75	0.0001	1366	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-548-608-19	+	156.00	198.75	0.0001	1366	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-316-650-12	+	155.00	196.71	0.0001	1442	

/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-056-556-112 + 155.00 196.71 0.0001 1213
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-413-814-79 - 153.50 196.37 0.0002 747
/cgn2_6/ptodata/2/1aa/6A_COMB.ppt:US-09-034-177-3 + 152.00 199.36 0.0002 1078
/cgn2_6/ptodata/2/1aa/6A_COMB.ppt:US-08-963-825-21 + 152.00 195.44 0.0002 1078
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-570-573-21 + 152.00 195.44 0.0002 1078
seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-112-114
seq_documentation_block:
; Sequence 114, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-114
alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
alignment_block:
US-09-462-480-1 x US-08-818-112-114 ..
Align seg 1/1 to: US-08-818-112-114 from: 1 to: 368.
1 CTCAGCAGGTGACGTGTTGTTTCAGCCAGGTGGCGGCGGCGCGGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
51 CAACCCAGCGCAGCAGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGG 100
|||||
242 YasnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCCCGCGGCGGCGG 150

US-08-818-111-109

APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998

```
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-114

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
  US-09-462-480-1 x US-09-056-556-114 ..
  Align seg 1/1 to: US-09-056-556-114 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGCTGCTTCAGCCAGGTGGCGCGCGCGGG 50
  |||||||
236 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
  |||||||

51 CAACCCAGCCGACGAGGAGCGCGGCGAGATGGCGCTGCTGCGCACGTC 100
  |||||||
242 YAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerp 259
  |||||||

101 CGCTGTGCAACATCCGCTGCTGCTGGATCAGGCCCGCGCGCGCG 150
  |||||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
  |||||||

151 GGCCTGTGCGCGCGGAGTCTACCTGCGCGAGGTGGTGGTGCACCG 200
  |||||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
  |||||||

201 CACGCCGTGATGTCTACGCTGATCGAAAGCCGCTGCCCGCTCGGTGA 250
  |||||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValaProSerValM 309
  |||||||

251 TCCCGCGCGGTGTCGCCGATCGTGGTGACGGGTGGCGCGCTCCGGTG 300
  |||||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
  |||||||

301 GGTCCGGGAGCGATGGCGCGGTTGCAATCCGCGCGCTCCACGCGCC 350
  |||||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
  |||||||

351 GGGTGTGTCGCGCGCGCACCGCTCGCGCAGGAGGTGAAGAAGACGACG 400
  |||||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
  |||||||

401 AGGACGACTGGACGAAGACGACGACTGG 429
  |||||||
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-056-556-214

seq_documentation_block:
; Sequence 214, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
```

TREATM

```
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-214

alignment_scores:
  Quality: 487.00      Length: 103
  Ratio: 4.822         Gaps: 0
  Percent Similarity: 98.058  Percent Identity: 97.087

alignment_block:
  US-09-462-480-1 x US-09-056-556-214 ..
  Align seg 1/1 to: US-09-056-556-214 from: 1 to: 802

522 AGCATGCGCAGATGAAGACCGATGCCGCTACCCCTCGGCGCAGGCGAG 571
  |||
423 SerSerAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGl 439
  |||||||

572 TAATTTGAGCGGATCTCCGCGACCTGAAACCCAGATCGACCGGTGG 621
  |||||||
439 yAsnPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValG 456
  |||||||

622 AGTCACGCGCAGGTTGTTGCGAGGCCAGTGGCGCGCGCGCGCGGAGG 671
  |||||||
456 luSerThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 472
  |||||||

672 GCGCGCCAGCGCGCGTGTGCTTCCAAAGACGAGCAATAAGCAGAA 721
  |||||||
473 AlaAlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnly 489
  |||||||

722 GCAGGAATCGACGAGATCTCGACGAATATTGTCAGGCCGCGCGTCCAA 771
  |||||||
489 sgGlnGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGln 506
  |||||||

772 ACTCAGGCGCGCAGGAGCAGCAGCAGCGCTGCTCTCGCAATGGCG 821
  |||||||
506 yrSerArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGly 522
  |||||||

822 TTCTGACCC 830
  |||
523 PheValPro 525

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-818-112-115
```

```
seq_documentation_block:
; Sequence 115, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-818-112-115

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-1 x US-08-818-112-115 ..
Align seg 1/1 to: US-08-818-112-115 from: 1 to: 100

525 ATGGCAGAGATGAGACCGGCTACCTCGCGCAGGAGGAGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17

575 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34

625 CGACGCGAGGTTCGTTGTCAGGCGCCAGTGGCGCGCGGGGACGGCC 674
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50

675 GCCACGCGCGGTGGTGGCTTCCAAAGCAGCAGCAATAAGCAGAACA 724
|||||
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67

725 GGAACCTCGACGAGATCTCGACGAATATTCGTGACGCGCGCGGTCCAATACT 774
|||||
```

```
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
|||||
775 CGAGGGCCCGACGAGGAGCAGCAGCGCTGTCTCGCAAAATGGGCTTC 824
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-818-111-110

seq_documentation_block:
; Sequence 110, Application US/08818111
; Patent No. 6338552
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-818-111-110

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-1 x US-08-818-111-110 ..
Align seg 1/1 to: US-08-818-111-110 from: 1 to: 100

525 ATGGCAGAGATGAGACCGGCTACCTCGCGCAGGAGGAGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17

575 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34

625 CGACGCGAGGTTCGTTGTCAGGCGCCAGTGGCGCGCGGGGACGGCC 674
|||||
```

|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
675 GCCAGCGCGGTTGCGCTTCCAGAGCAGCCCAATAGCAGACGA 724
|||||
51 AlaGlnAlaAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
725 GGAACCTGACGAGATCTCGACGAATATTCTGTCAGCGCGCTCCAACTACT 774
|||||
67 nGluLeuAspGluLeuSerThrAsnIleArgGlnAlaGlyValGlnTrp 84
775 CGAGGCGCGAGCAGAGCAGCAGCGCTGCTCTCCGAATAGGCTTC 824
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-056-556-115

seq_documentation_block:
; Sequence 115, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Asir A.W.
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-056-556-115

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:

US-09-462-480-1 x US-09-056-556-115 ..

Align seg 1/1 to: US-09-056-556-115 from: 1 to: 100

525 ATGGCAGAGATCAAGACCGATCGCGTACCCTCGGCGAGGAGCGGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
575 TTTCAGCGGATCTCCGCGCGACCTGAAACCCAGATCGACCGGTGGAGT 624

|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
625 CGACGCGAGGTTCCTTCAGGCGCAGTGGCGCGCGGGGGGAGCGGCC 674
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
675 GCCAGCGCGGTTGCGCTTCCAGAGCAGCCCAATAGCAGACGA 724
|||||
51 AlaGlnAlaAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
725 GGAACCTGACGAGATCTCGACGAATATTCTGTCAGCGCGCTCCAACTACT 774
|||||
67 nGluLeuAspGluLeuSerThrAsnIleArgGlnAlaGlyValGlnTrp 84
775 CGAGGCGCGAGCAGAGCAGCAGCGCTGCTCTCCGAATAGGCTTC 824
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-465-640-2

seq_documentation_block:
; Sequence 2, Application US/08465640
; Patent No. 5955077
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: ANDERSEN, Ase Bengaard
; APPLICANT: HASLOV, Kaare
; APPLICANT: SORENSEN, Anne Lund
; TITLE OF INVENTION: TUBERCULOSIS VACCINE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,640
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,182
; FILING DATE: 20-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK94/00273
; FILING DATE: 01-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: ANDERSEN-3A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-465-640-2

alignment_scores:
Quality: 479.00 Length: 95
Ratio: 5.042 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
US-09-462-480-1 x US-08-465-640-2
Align seg 1/1 to: US-08-465-640-2 from: 1 to: 95

860 ATGACAGACGACGTGGAAATTCGCGGTATCGAGCGCGGCAAGCGC 909
|||||
1 MetThrGluGlnGlnTrpAsnPheAlaGlyIleGluAlaAlaSerAl 17

910 AATCCAGGAATGTCAGCTCATTCCTCTTCACGAGGGAAGC 959
|||||
17 aileGlnGlyAsnValThrSerIleHisSerLeuLeuAspGluGlyIysG 34

960 AGTCCCTGACCAAGCTCGACGCGCTGGGCGGTAGCGTTTCGAGGCG 1009
|||||
34 InSerLeuThrLysLeuAlaAlaAlaIleArgGlySerGlySerGluAla 50

1010 TACCAGGGGTGCCAGCAAAATGGACGCCACCGCTACCGAGCTGAACAA 1059
|||||
51 TyrGlnGlyValGlnGlnLysTrpAspAlaThrAlaThrGluLeuAsnAs 67

1060 CGCGCTGACGAACCTGGCGCGGAGATCAGCAAGCGCGGTTCAGGCAATGG 1109
|||||
67 nAlaLeuGlnAsnLeuAlaArgThrIleSerGluAlaGlyGlnAlaMeta 84

1110 CTCGACCGAAGCAACGTCACCTGCGATGTCGCA 1144
|||||
84 laSerThrGluGlyAsnValThrGlyMetPheAla 95

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-818-112-88

seq_documentation_block:
; Sequence 88, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS.
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-88

alignment_scores:
Quality: 462.00 Length: 95
Ratio: 4.863 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.947

alignment_block:
US-09-462-480-1 x US-08-818-112-88
Align seg 1/1 to: US-08-818-112-88 from: 1 to: 95

540 ACCGATCCCGCTACCCCTCGGCGAGGAGCAGTAAATTCGAGCGGATCTC 589
|||||
1 ThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPheGluArgIleSe 17

590 CGCGACCTGAAACCCAGATCGACCGAGTGGAGTCGACGCGCAGGTTCTG 639
|||||
17 rGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaGlySerL 34

640 TGCAGGCGCAGTGGCGCGCGCGGCGGCGCGCGCGCGCGCGCGGTG 689
|||||
34 euGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAlaAlaVal 50

690 GTGCGCTTCCAAAGAGCAGCCCAATAAGCAGAAAGCAGGAACTCGAGGAT 739
|||||
51 ValArgPheGlnGlnAlaAlaAsnLysGlnLysGlnGluLeuAspGluIl 67

740 CTCGACGAATATTCGTCAGGCGCGCTCTCTCGCAATCTCGAGGCGCGAGG 789
|||||
67 eSerThrAsnIleargGlnAlaGlyValGlnTrpSerArgAlaAspGluG 84

790 AGCAGCAGCAGCGCGCTCTCTCGCAATCTCGCAATCTCGCGTTC 824
|||||
84 luGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 95

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-818-111-89

seq_documentation_block:
; Sequence 89, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
```



```

; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; NAME: Maki, David J.
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-89

alignment_scores:
    Quality: 462.00      Length: 95
    Ratio: 4.863        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.947

alignment_block:
US-09-462-480-1 x US-08-818-111-89 ..
Align seg 1/1 to: US-08-818-111-89 from: 1 to: 95

540 ACCGATGCCGTACCTCGGCAGGAGCAGGTAATTTCGAGCGGATCTC 589
|||||
1 ThrAspAlaAThrLeuAlaGlnGluAlaGlyAsnPheGluArgIleSe 17

590 CGGCGACTGAAACCAGATCGACCCAGGTGGAGTTCGACGCGGTTGCT 639
|||||
17 rGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaGlySerL 34

640 TGCAGGGCCAGTGCGCGCGCGCGGGAGGAGCGCCGCCAGCGCGGTG 689
|||||
34 euGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAlaAlaVal 50

690 GTGCGCTTCCAAGAAGCAGCAATAAGCAGAGGAGGAGGAGGAGGAGG 739
|||||
51 ValArgPheGlnGluAlaAlaAsnLysGlnLysGlnGluLeuAspGluIl 67

740 CTCGACGAATATTCGTGAGCGCGCGTCCCAATCTCGAGGCGCGCTTC 824
|||||
84 luGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 95

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-056-556-88

seq_documentation_block:
; Sequence 88, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998

```

ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.41106
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 117:
SEQUENCE CHARACTERISTICS:
LENGTH: 80 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-117

alignment_scores:
Quality: 392.00 Length: 80
Ratio: 4.900 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-462-480-1 x US-08-818-112-117

Align seg 1/1 to: US-08-818-112-117 from: 1 to: 80

585 ATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGTCGACGCGAGG 634
|||||
1 lIleSerGlyAspLeuylThrGlnIleaspGlnValGluSerThrAlaGl 17
|||||
635 TTCGTTGCGAGGCCAGTGGCGCGCGCGGCGGACGCGCGCCCGCCAGGCGG 684
|||||
17 ySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAla 34
|||||
685 CGGTGGTGGCTTCCAAAGCAGCCATAGCAGAACGAGAACTCGAC 734
|||||
34 lAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysGlnLysGlnLeuAsp 50
|||||
735 GAGATCTCGAGCAATATTTCGTGCGCGCGCGCTCCCAATAGTGGCTTC 824
|||||
51 GlulSerThrAsnIleArgGlnAlaGlyValGlnIleGlnIleGlnIleGlnIle 67
|||||
785 CGAGAGCAGCAGCGCGCTGTCTCTCGCAAAATGGGCTTC 824
|||||
67 pGluGluGlnGlnGlnAlaLeuSerSerGlnMetClyPhe 80
|||||

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-818-111-112

seq_documentation_block:

Sequence 112, Application US/08818111

Patent No. 6338852

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonia

APPLICANT: Houghton, Raymond

APPLICANT: Vedvick, Thomas S.

APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

TUBERCULOSIS

NUMBER OF SEQUENCES: 148

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.41706
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 112:
SEQUENCE CHARACTERISTICS:
LENGTH: 80 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-112

alignment_scores:

Quality: 392.00 Length: 80
Ratio: 4.900 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-462-480-1 x US-08-818-111-112

Align seg 1/1 to: US-08-818-111-112 from: 1 to: 80

585 ATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGTCGACGCGAGG 634
|||||
1 lIleSerGlyAspLeuylThrGlnIleaspGlnValGluSerThrAlaGl 17
|||||
635 TTCGTTGCGAGGCCAGTGGCGCGCGCGGCGGACGCGCGCCCGCCAGGCGG 684
|||||
17 ySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAla 34
|||||
685 CGGTGGTGGCTTCCAAAGCAGCCATAGCAGAACGAGAACTCGAC 734
|||||
34 lAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysGlnLysGlnLeuAsp 50
|||||
735 GAGATCTCGAGCAATATTTCGTGCGCGCGCGCTCCCAATAGTGGCTTC 824
|||||
51 GlulSerThrAsnIleArgGlnAlaGlyValGlnIleGlnIleGlnIleGlnIle 67
|||||
785 CGAGAGCAGCAGCGCGCTGTCTCTCGCAAAATGGGCTTC 824
|||||
67 pGluGluGlnGlnGlnAlaLeuSerSerGlnMetClyPhe 80
|||||

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-056-556-117

seq_documentation_block:

Sequence 117, Application US/09056556

Patent No. 6350456

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND

TUBERCULOSIS

NUMBER OF SEQUENCES: 241

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mak1, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 117:
SEQUENCE CHARACTERISTICS:
LENGTH: 80 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-117

alignment_scores:
Quality: 392.00 Length: 80
Ratio: 4.900 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-480-1 x US-09-056-556-117 ..

Align seg 1/1 to: US-09-056-556-117 from: 1 to: 80

585 ATCTCCGGCGACCTCAAAACCCAGATCCAGGTCAGGTCGACGGCAGG 634
|||||
1 IleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaG1 17
635 TTCGTTGACGGCCAGTCGCGCGCGCGGACGCCGCCCGCCAGCGG 684
|||||
17 ySerLeuGlnGlyGlnTrpArgGlyAlaGlyThrAlaAlaGlnAlaA 34
685 CGGTGGTGGCTTCCAAAGACGAGCCAAATAGCAGACGAGGAAGTCGAC 734
|||||
34 laValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnGluLeuAsp 50
735 GAGATCTCGAGCAATATTCGTACGCGCGCGCTCCCAATACTCGAGGCGGA 784
|||||
51 GluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAs 67
785 CGAGGAGCAGCAGCAGCGGCTGCTCTCCCAATGGCTTC 824
|||||
67 pGluGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 80

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-818-112-104

seq_documentation_block:

Sequence 104, Application US/08818112
Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Mak1, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-104

alignment_scores:
Quality: 259.00 Length: 51
Ratio: 5.078 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-480-1 x US-08-818-112-104 ..

Align seg 1/1 to: US-08-818-112-104 from: 1 to: 51

860 ATGACAGACGACGACTGGAATTCGCGGTCATCGAGCGCGCAAGCGC 909
|||||
1 MetThrGluGlnGlnTrpAsnPheAlaGlyIleGluAlaAlaAlaSerAl 17
910 AATCCAGGGAATGTACGTCCATTCTCCCTCTGACGAGGGAAGC 959
|||||
17 alIeGlnGlyAsnValThrSerIleHisSerLeuLeuAspGluGlyLysG 34
960 AGTCCTGACCAAGCTCCGACGGCGCTGGGCGGTAGCGGTTCGAGCGG 1009
|||||
34 InSerLeuThrLysLeuAlaAlaAlaTrpGlyGlySerGlySerGluAla 50
1010 TAC 1012
|||
51 Tyr 51

OM of: US-09-462-480-1 to: Pending_Patents_AA_Main:* out_format : pfs
Date: Jul 22, 2002 1:35 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+nt2p_model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09462480/runat_18072002.164419_19468/app_query.fasta_1.2850
-DB=Pending_Patents_AA_Main -OFMT=fastan -SUFFIX=rapm
-GAPO=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -FGAPO=4.500 -FGAPEXT=0.050 -XGAPO=10.000
-XGAPEXT=0.500 -FGAPO=6.000 -FGAPEXT=7.000 -YGAPO=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-NATRIX=blosum62 -TRANS=human40 cdi -LIST=45 -DOCLIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09462480 -CGN1_1_271 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-1
Query length: 1277
Database: Pending_Patents_AA_Main:*
Database sequences: 3502263
Database length: 351980561
Search time (sec): 542.380000

score_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03265-109		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03268-114		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-658-800-109		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-659-683-114		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-680-573-109		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-680-574-114		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-620-109		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-720-622-109		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-730-510-114		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-341-109		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-578-114		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-024-753-109		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-025-197-114		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-596-109		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-967-114		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-724-685-114		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-724-685-114		713.00	813.95	3.0e-37	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-050-739-92		713.00	813.87	3.0e-37	371
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-246-191-92		713.00	813.87	3.0e-37	371
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-791-171-92		713.00	813.87	3.0e-37	371
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-804-980-92		713.00	813.87	3.0e-37	371
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US-09-943-443-92		713.00	813.87	3.0e-37	371
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-263-688-2		512.50	589.88	1.7e-24	196
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-287-849-8		494.00	563.00	2.9e-23	358
/cgn2_6/ptodata/2/paa/US094_COMB.ppt:US-09-116-492-5		492.00	572.69	3.0e-23	100
/cgn2_6/ptodata/2/paa/US094_COMB.ppt:US-09-462-480-5		492.00	572.69	3.0e-23	100
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-050-733-172		487.50	554.41	7.7e-23	404
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-246-191-172		487.50	554.41	7.7e-23	404
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-791-171-172		487.50	554.41	7.7e-23	404
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-804-980-172		487.50	554.41	7.7e-23	404
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-805-427A-3		487.50	554.41	7.7e-23	404
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US-09-943-443-172		487.50	554.41	7.7e-23	404
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03265-209		487.00	547.39	9.6e-23	802
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03265-346		487.00	547.39	9.6e-23	802
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03268-214		487.00	547.39	9.6e-23	802
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03268-351		487.00	547.39	9.6e-23	802
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-341-209		487.00	547.39	9.6e-23	802
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-578-214		487.00	547.39	9.6e-23	802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-024-753-209		487.00	547.39	9.6e-23	802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-025-197-214		487.00	547.39	9.6e-23	802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-596-209		487.00	547.39	9.6e-23	802

/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-596-346 + 487.00 547.39 9.6e-23
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-967-214 + 487.00 547.39 9.6e-23
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-967-351 + 487.00 547.39 9.6e-23
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-287-849-10 + 487.00 547.39 9.6e-23
/cgn2_6/ptodata/2/paa/US100_COMB.ppt:US-10-084-843-214 + 487.00 547.39 9.6e-23
seq_name: /cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03265-109
seq_documentation_block:
; Sequence 109, Application PC/TUS9903265
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US99/03265
; FILING DATE: 17-FEB-1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/024,753
; FILING DATE: 18-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9532-0023-228
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US99-03265-109
alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
alignment_block:
US-09-462-480-1 x PCT-US99-03265-109
Align seg 1/1 to: PCT-US99-03265-109 from: 1 to: 368
1 CTGACGAGGTGACGCTGTTGTTTTCAGGCAGGTGGCGGCGACCGCGCGGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG1 242
51 CAACCCAGCCGACGAGGAGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCCGAACATCCGCTGCTGCTGGATCAGCCGCGGCGGCGGCGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

```
151 GGCCTGCTGCGCGGAGTCCCTACCTGGCGCAGGTGGTGGTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACCGCGTGTGCTCAGCTGATCGAAAGCCGCTTCCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGGAGTGGCGGAGGCTGCGATCCGATCCGCGGCTCCACGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspaspG 359
|||||
401 AGGACGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:PCT-US99-03268-114

```
seq_documentation_block:
; Sequence 114, Application PC/TUS9903268
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US99/03268
; FILING DATE: 17-FEB-1999
; CLASSIFICATION: 406
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/025,197
; FILING DATE: 18-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9532-0013-228
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
PCT-US99-03268-114
```

```
alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
```

```
alignment_block:
US-09-462-480-1 x PCT-US99-03268-114
Align seg 1/1 to: PCT-US99-03268-114 from: 1 to: 368
1 CTGACGAGGTGACGTGCTGTTCTTCCAGCCAGGTGGCGGCACCCGGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
|||||
51 CAACCCGCCCGACGAGGAAGCCGCGCAGATGGCGCTGCTCGGCACCGATC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGCTGCTGATCAGGCCGCCAGCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGCTGCGCGGCGGAGTCCCTACCTGGCGCAGGTGGTGGTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACCGCGTGTGCTCAGCTGATCGAAAGCCGCTTCCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGGAGTGGCGGAGGCTGCGATCCGATCCGCGGCTCCACGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspaspG 359
|||||
401 AGGACGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US086_COMB.pep:US-08-658-800-109

```
seq_documentation_block:
; Sequence 109, Application US/08658800
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,800
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C3
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 368 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;
US-08-658-800-109

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
US-09-462-480-1 x US-08-658-800-109 ..
Align seg 1/1 to: US-08-658-800-109 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGTTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242

51 CAACCCAGCGGACGAGGAGCGCGCAGATGGGCTGCTGGCACCAAGTC 100
|||||
242 YAsnProAlaAspGluAlaGlnMetGlyLeuGlyThrSerP 259

101 CGGTGTCGAACCATCCGCTGGCTGGTGCATCAGCCCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCTGTGTCGCGCGGAGTCGCTACCTGCGCGCAGGTGGTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292

201 CACGCCGCTGATGTCTCAGCTGATCGAAAGCCGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValm 309

251 TGCCGGCGGCTGTTCCGGATCGCTGGTGCAGGCGGCGCGCGCGTGC 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325

301 GGTCCGGAGGATGGCGCGGTCGCAATCCGCGCGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342

351 GGGTCTGGTCGCGCGCACCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359

401 AGGACGACTGGGACGAGACGACACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US086_COMB.pep:US-08-659-683-114

seq_documentation_block:
; Sequence 114, Application US/08659683
; GENERAL INFORMATION:
;   APPLICANT: Reed, Steven G.
;   APPLICANT: Skeiky, Yasir A.W.
;   APPLICANT: Dillon, Davin C.
;   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
;   TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
;   NUMBER OF SEQUENCES: 128
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SEED AND BERRY LLP
;   STREET: 6300 Columbia Center, 701 Fifth Avenue
;   CITY: Seattle
;   STATE: Washington
```

```
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,683
; FILING DATE: 05-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Maki, David J.
;   REGISTRATION NUMBER: 31,392
;   REFERENCE/DOCKET NUMBER: 210121.411C3
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 622-4900
;   TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 368 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;
US-08-659-683-114

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
US-09-462-480-1 x US-08-659-683-114 ..
Align seg 1/1 to: US-08-659-683-114 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGTTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242

51 CAACCCAGCGGACGAGGAGCGCGCAGATGGGCTGCTGGCACCAAGTC 100
|||||
242 YAsnProAlaAspGluAlaGlnMetGlyLeuGlyThrSerP 259

101 CGGTGTCGAACCATCCGCTGGCTGGTGCATCAGCCCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCTGTGTCGCGCGGAGTCGCTACCTGCGCGCAGGTGGTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292

201 CACGCCGCTGATGTCTCAGCTGATCGAAAGCCGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValm 309

251 TGCCGGCGGCTGTTCCGGATCGCTGGTGCAGGCGGCGCGCGCGTGC 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325

301 GGTCCGGAGGATGGCGCGGTCGCAATCCGCGCGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342

351 GGGTCTGGTCGCGCGCACCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359

401 AGGACGACTGGGACGAGACGACACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq_name: /cgn2_6/ptodata/2/paa/US086_COMB.pep.US-08-680-573-109

seq_documentation_block:
; Sequence 109, Application US/08680573
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,573
; FILING DATE: 12-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-680-573-109

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-1 x US-08-680-573-109 ..
Align seg 1/1 to: US-08-680-573-109 from: 1 to: 368

```
1 CTCGACGAGTGAGTGGTGTTCAGCCAGGTGGCGGCGCACCGCGG 50
|||||
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
|||||
51 CAACCCAGCCGACGAGGAAGCGCGGAGATGGCTCTCGGCACCAATC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuGlyThrSerP 259
|||||
101 CGCTGTCAACACCATCCGCTGGCTGGTGGATCAGGCCCGCCAGCGGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGTGTCGCGCGGAGTGCCTACCTGGCGGCGAGGTGGTGTGACCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThr 292
|||||
201 CAGCCCGCTGATGCTCAGCTGATCGAAAGCGGTTGCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309
|||||
```

```
251 TGCCGGCGGCTGTTCGCGGATCGTGGTGACGGGTGGCGGCTCCGCGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGGATGGCGGAGGTTCGAATCCGCGGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTCGCGCGGCGCACCGCTCGGCGAGGAGCGTGAAGAAGACGAG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US086_COMB.pep.US-08-680-574-114

seq_documentation_block:
; Sequence 114, Application US/08680574
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 133
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,574
; FILING DATE: 12-JUL-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-680-574-114

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-1 x US-08-680-574-114 ..
Align seg 1/1 to: US-08-680-574-114 from: 1 to: 368

```
1 CTCGACGAGTGAGTGGTGTTCAGCCAGGTGGCGGCGCACCGCGG 50
|||||
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
|||||
```



```
51 CAACCCAGCGAGCAGGAGCGCGCAGATGGCGCTGCTCGGCACCAAGTC 100
|||||
242 YasnProAlaAspGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCGAACCATCCGCTGGCTGGTGATCAGGCCCGCCAGCGGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGCTGCGCGCGGAGTCCCTACCTGCGCGCAGGTGGCTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCGGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309
|||||
251 TGGCGGGCGCTGTTGCGGGATGCTGCTGACGGTGCGCGCTCGCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaPro 325
|||||
301 GGTGCTGCTGCGCGCGCAGGTTCGCAATCCGGCGGCTCCACGACCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
|||||
351 GGTGCTGCTGCGCGCGCAGGTTCGCGCAGGAGCGGTGAAGAACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
|||||
401 AGGACGACTGGGAGGAGGAGGAGGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US087_COMB.pap:US-08-729-622-109

```
seq_documentation_block:
; Sequence 109, Application US/08729622
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas H.
; APPLICANT: Twardzik, David R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 132
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/729,622
; FILING DATE: 11-OCT-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C5
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-729-622-109

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-1 x US-08-729-622-109 ..
Align seg 1/1 to: US-08-729-622-109 from: 1 to: 368

1 CTGACGAGGTGAGCTGCTGTTCTCAGCCAGGTGGCGCGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
|||||
51 CAACCCAGCGAGCAGGAGCGCGCAGATGGCGCTGCTCGGCACCAAGTC 100
|||||
242 YasnProAlaAspGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCGAACCATCCGCTGGCTGGTGATCAGGCCCGCCAGCGGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGCTGCGCGCGGAGTCCCTACCTGCGCGCAGGTGGCTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCGGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309
|||||
251 TGGCGGGCGCTGTTGCGGGATGCTGCTGACGGTGCGCGCTCGCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaPro 325
|||||
301 GGTGCTGCTGCGCGCGCAGGTTCGCAATCCGGCGGCTCCACGACCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
|||||
351 GGTGCTGCTGCGCGCGCAGGTTCGCGCAGGAGCGGTGAAGAACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
|||||
401 AGGACGACTGGGAGGAGGAGGAGGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US087_COMB.pap:US-08-730-510-114

```
seq_documentation_block:
; Sequence 114, Application US/08730510
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 137
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
```

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/730,510
FILING DATE: 27-AUG-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.41105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 368 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-730-510-114

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-1 x US-08-730-510-114

Align seg 1/1 to: US-08-730-510-114 from: 1 to: 368

1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCGACGAGGAGCGCGCAGATGGCTCTCGGCACCACTC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCTGAACCATCCGCTGCTGGTGGATCAGGCCCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTCTGCGCGGAGTCTGCTACCTGGCGCAGTGGGTGTTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCGGTGATGCTCAGCTGATCGAAAGCCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValm 309
251 TGCCGCGGCTGTTGCCGATCTGCTGGTGGTGGCGCGCTCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGAGCGATGGCGAGGTTTCGAATCCGCGCGCTCCACACGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlySerThrArgr 342
351 GGGTCTGTGTCGCGCGCACCGCTCGCGCAGGAGCGGTGAAGAACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspasp 359
401 AGNACGACTGGGACGAAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US089_COMB.pap:US-08-942-341-109

seq_documentation_block:

: Sequence 109, Application US/08942341

: GENERAL INFORMATION:

: APPLICANT: Reed, Steven G.

: APPLICANT: Skeiky, Yasir A.W.

: APPLICANT: Dillon, Davin C.

: APPLICANT: Campos-Neto, Antonia

: APPLICANT: Houghton, Raymond

: APPLICANT: Vedvick, Thomas S.

: APPLICANT: Twardzik, Daniel R.

: APPLICANT: Loges, Michael J.

: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

: NUMBER OF SEQUENCES: 209

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: SEED and BERRY LLP

: STREET: 6300 Columbia Center, 701 Fifth Avenue

: CITY: Seattle

: STATE: Washington

: COUNTRY: USA

: ZIP: 98104-7092

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/942,341

: FILING DATE: 01-OCT-1997

: CLASSIFICATION: 435

: ATTORNEY/AGENT INFORMATION:

: NAME: Maki, David J.

: REGISTRATION NUMBER: 31,392

: REFERENCE/DOCKET NUMBER: 210121.417C7

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (206) 622-4900

: TELEFAX: (206) 682-6031

: INFORMATION FOR SEQ ID NO: 109:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 368 amino acids

: TYPE: amino acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: US-08-942-341-109

alignment_scores:

Quality: 713.00 Length: 143

Ratio: 5.057 Gaps: 0

Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-1 x US-08-942-341-109

Align seg 1/1 to: US-08-942-341-109 from: 1 to: 368

1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGTGGCGGCACCGCGCGG 50

|||||

226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242

51 CAACCCAGCGACGAGGAGCGCGCAGATGGCTCTCGGCACCACTC 100

|||||

242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTCTGAACCATCCGCTGCTGGTGGATCAGGCCCGCGCGCG 150

|||||

259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCCTGTCTGCGCGGAGTCTGCTACCTGGCGCAGTGGGTGTTGACCCG 200

|||||

276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292

```
201 CACCCCGCTGATGCTCAGCTGATCGAAGAACCGGTTGCCCGCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TCCCGCGCGGTGTTGCCGGATCGCTCGGTGACGGGTGCGCGCTCCGGTG 300
309 etProAlaAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGAGCGATGGCGGAGGTTCCAAATCCGGCGGCTCCACGACGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGTCGCGCGGACCGCTCGCGCAGGAGCGTGAAGAACGACGACG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGACGACTGGGACGAGAGGACGACTGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq_name: /cgn2_6/ptodata/2/paa/US089_COMB.pep:US-08-942-578-114

```
seq_documentation_block:
; Sequence 114, Application US/08942578
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/942,578
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-942-578-114
```

alignment_scores: Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

```
alignment_block:
US-09-462-480-1 x US-08-942-578-114
Align seg 1/1 to: US-08-942-578-114 from: 1 to: 368
1 CTGCAGCAGGTGACGTGCTGTTTCAGCAGGTGGCGGACCGCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyGly 242
51 CACCCAGCGCAGCAGGAGCCGCGGACAGATGGCGCTGCTCGGCACGATC 100
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGTGTGTCGAACCATCCGCTGCTGGATCAGGCCCGCCAGCGCGGCGCG 150
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTGCGCGCGGAGTGCCTACCTGGCGCAGGTGGGTGCTTGACCCG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACCCCGCTGATGCTCAGCTGATCGAAGAACCGGTTGCCCGCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TCCCGCGCGGTGTTGCCGGATCGCTCGGTGACGGGTGCGCGCTCCGGTG 300
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGAGCGATGGCGGAGGTTCCAAATCCGGCGGCTCCACGACGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGTCTGTCGCGCGGACCGCTCGCGCAGGAGCGTGAAGAACGACGACG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGACGACTGGGACGAGAGGACGACTGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-024-753-109

```
seq_documentation_block:
; Sequence 109, Application US/09024753
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 236
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/024,753
; FILING DATE: 18-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
```

TUBERCULOSIS

```

; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-024-753-109

alignment_scores:
    Quality: 713.00      Length: 143
    Ratio: 5.057        Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-1 x US-09-024-753-109

Align seg 1/1 to: US-09-024-753-109 from: 1 to: 368

1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAGCGGCGGAGTGGCTGCTCGGCACCGATC 100
|||||
242 YasnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGTAACATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCTGTGTCGACCATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTGATCGAATCCGCGGCTCCACCGCC 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerVal 309
251 TGCCGCGGCTGTTGCGCGGATCGTGGTGGTGGTGGTGGTGGTGGTGG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaPro 325
301 GGTCGCTGTCGCGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGG 350
|||||
326 GlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
401 AGGACGACTGGGCGAGAGGACGACTGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-025-197-114

seq_documentation_block:
; Sequence 114, Application US/09025197
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
```

351 GGGTCTGGTCCGCGCACCGCTCGCGCAGGAGCGTGAAGAACGACGACG 400
|||||
342 dGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSPG 359
|||||
401 AGGACGACTGGGACGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-072-596-109

seq_documentation_block:
; Sequence 109, Application US/09072596
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-109

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
alignment_block:
US-09-462-480-1 x US-09-072-596-109
Align seg 1/1 to: US-09-072-596-109 from: 1 to: 368
1 CTGACGAGGTGACGTCTGTTTCAGCCAGGTGGCGCAGCCGCGGCGG 50
|||||
236 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
|||||
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGCTGCTGCGCACCACTC 100
|||||

242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCTGAACCATCCGCTGGTGGATCAGGCCAGCCCGCGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGTCTGCGCGGAGTCTACCTGGCGCAGGTGGTGGTGGTGGTGA 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCCTGATGTCTCAGCTGATCGAAAGCCCGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309
|||||
251 TGCGCGCGGTGTTGGCGGATCGCTGCGTGGCGGCGGTGGCGGTG 300
|||||
309 etProAlaAlaAlaMetGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGCGATGGCCAGGGTTCGCAATCCGCGCGCTCCACCGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGTCTGTGTCGCGCGCGCGCGCTCGCGCAGGAGCGGTGAAGAACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSPG 359
|||||
401 AGGACGACTGGGACGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||
seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-072-967-114
seq_documentation_block:
; Sequence 114, Application US/09072967
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-967-114

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-1 x US-09-072-967-114

Align seg 1/1 to: US-09-072-967-114 from: 1 to: 368.

```
1 CTGCAGCAGGTGACGTCGTTGTTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAAGCGCGCAGATGGGCTGCTCGGCACCGAGTC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCCCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTGCGCGGGAGTCGCTACCTGGCGCAGGTGGGTGCTTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CACGCGCTGATGTCAGCTGATCGAAAGCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309
|||||
251 TGCCGGCGGCGTGTTCGCGAGTCGTCGTTGACGGGTGGCGCGCGCTCG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGGAGGATGGCGCAGGTTCGCAATCCGCGCGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
|||||
351 GGGTCTGTGTCGCGCGGCGCACCGCTCGCGCAGGAGCGTGAAGAAGC 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
```

seq_name: /cqn2_5/ptodata/2/paa/US097_COMB.pep:US-09-724-685-114

seq_documentation_block:

Sequence 114, Application US/09724685

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

Skeiky, Yasir

Dillon, Davin C.

Campos-Neto, Antonio

TITLE OF INVENTION: Compounds and Methods for

Immunotherapy and Diagnosis of Tuberculosis

NUMBER OF SEQUENCES: 155

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/724,685
FILING DATE: 28-Nov-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/523,436
FILING DATE: 01-SEP-1995
APPLICATION NUMBER: US 08/533,634
FILING DATE: 22-SEP-1995
APPLICATION NUMBER: US 08/620,874
FILING DATE: 22-MAR-1996
APPLICATION NUMBER: US 08/659,683
FILING DATE: 05-JUN-1996
APPLICATION NUMBER: US 08/680,574
FILING DATE: 12-JUL-1996
APPLICATION NUMBER: WO PCT/US96/14674
FILING DATE: 30-AUG-1996
APPLICATION NUMBER: US 08/730,511
FILING DATE: 11-OCT-1996
APPLICATION NUMBER: US 08/818,112
FILING DATE: 13-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 014058-008561US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 368 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 114:
US-09-724-685-114

alignment_scores:

Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:

US-09-462-480-1 x US-09-724-685-114

Align seg 1/1 to: US-09-724-685-114 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTCGTTGTTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAAGCGCGCAGATGGGCTGCTCGGCACCGAGTC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCCCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTGCGCGGGAGTCGCTACCTGGCGCAGGTGGGTGCTTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CACGCGCTGATGTCAGCTGATCGAAAGCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309
|||||
251 TGCCGGCGGCGTGTTCGCGAGTCGTCGTTGACGGGTGGCGCGCGCTCG 300
|||||
```


THIS PAGE BLANK (USPTO)

OM of: US-09-462-480-1 to: Pending_Patents_AA_New:* out_format : pfs
Date: Jul 22, 2002 1:37 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2_1/uspt0.spool/US09462480/runat_18072002.164419_19498/app_query.fasta_1.2850
-DB=Pending_Patents_AA_New -QFMT=fastn -SUFFIX=rapn
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosom62 -TRANS=human40.cdi -LIST=45 -DICALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=LOCAL
-OUTFMT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09462480 -CGN1_1_93 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-1
Query length: 1277
Database: Pending_Patents_AA_New:*
Database sequences: 378952
Database length: 124292526
Search time (sec): 104.950000

score_list:

Sequence	Strd Orig	zscore	Escore	Len	Documentation
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-5		492.00	493.20	2.9e-19	10
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-5		492.00	493.20	2.9e-19	100
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-10-080-170B-639		486.00	487.37	6.0e-19	49
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-6		247.00	260.53	5.3e-06	49
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-6		247.00	260.53	5.3e-06	49
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17077		210.50	202.33	0.0006	49
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17076		210.50	202.13	0.0006	49
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17480		210.50	202.02	0.0006	49
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17479		210.50	201.84	0.0006	49
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17478		210.50	201.73	0.0006	49
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-7		205.00	201.62	0.0006	42
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-7		205.00	220.89	0.0010	42
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2231		203.50	187.43	0.0014	42
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-13		198.00	198.25	0.0025	3
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10704		191.50	190.60	0.0058	3
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-28		187.00	196.41	0.0097	10
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-28		187.00	196.41	0.0097	10
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-12		187.00	196.41	0.0097	10
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-11		186.00	195.85	0.0110	9
/cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:PCT-US02-13142-3473		181.50	170.81	0.0208	9
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-128-714-3473		181.50	170.81	0.0208	9
/cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:PCT-US02-13142-8473		175.50	164.87	0.0440	9
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-128-714-8473		175.50	164.87	0.0440	9
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-26077		172.50	171.65	0.0617	9
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-26078		171.00	166.24	0.0755	9
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-667-170A-187		167.00	162.64	0.1242	5
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-365-264-664		166.50	162.80	0.1319	5
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-360-039-20627		165.00	158.96	0.1605	5
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2448		165.00	156.09	0.1622	5
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-582-687A-60		164.50	154.68	0.1732	1
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17480		164.00	156.77	0.1826	1
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17479		164.00	156.59	0.1828	1
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-11792		163.50	163.09	0.1896	1
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-121-988-47		163.50	156.60	0.1942	82
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2272		162.00	152.22	0.2367	82
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2617		161.00	158.44	0.2611	82
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10213		161.00	156.95	0.2626	82
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-360-039-1813		161.00	154.26	0.2652	82

/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2231 - 159.50 144.62 0.3295
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-994-404-252 - 159.00 141.73 0.3538
/cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:PCT-US02-19669-70 + 158.50 146.97 0.3688
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-9406 - 158.50 146.48 0.3694
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-26246 - 158.50 146.48 0.3694
seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-5
seq_documentation_block:
; Sequence 5, Application US/09116492A
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM M
; TITLE OF INVENTION: TUBERCULOSIS ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS,
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492A
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-116-492A-5

alignment_scores:
Quality: 492.00 Length: 100
Ratio: 4.920 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-462-480-1 x US-09-116-492A-5 ..
Align seg 1/1 to: US-09-116-492A-5 from: 1 to: 100
525 ATGCAGAGATGAACCGATCGCGCTACCTCGGCGAGGAGGAGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlyGlnGluAlaGlyAs 17
575 TTTCAGCGGATCTCCGCGACCTGAAACCCAGATCGACGAGTGGAGT 624
|||||
17 nPheGluArgGlySerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
625 CGACGCGAGGTTCGTTTCAGGCGCAGTGGCGCGCGCGGCGGACGCC 674
|||||
34 erThrAlaGlySerLeuGlnGlyGlnIleArgGlyAlaAlaGlyThrAla 50
675 GCCAGCGCGCGGTGGTGGCGCTTCCAAAGAACGACGCCAATAAGCAAGCA 724
|||||
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67
725 GGAAGTCTGACGAGATCTCGAGCAATATTCGTCAGCGCGCGTCCAACT 774
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnIle 84
775 CGAGGCGCGAGGAGGACGACGAGCGGCTGTCTCGCAATGGCTTC 824
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe 100
seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-5
seq_documentation_block:
; Sequence 5, Application US/10140045
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER

```
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS A TOOL FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/10/140,045
; PRIOR FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: US/09/116,492A
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-140-045-5

alignment_scores:
  Quality: 492.00      Length: 100
  Ratio: 4.920        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-462-480-1 x US-10-140-045-5 ..
Align seg 1/1 to: US-10-140-045-5 from: 1 to: 100

525 ATGCAGAGATGAAGACCGATGCGCTACCTCGGCGAGGAGCGGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlnGluAlaGlyAs 17

575 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGGTGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34

625 CGAGCGCGGTTCGTTGCGAGCGGCGGTGCGGCGGCGGCGGCGGCC 674
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50

675 GCCAGCGCGGTTCGTTGCGAGCGGCGGTGCGGCGGCGGCGGCGGCC 724
|||||
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67

725 GGAACCTCGACGATCTCGACGATATTCGTCAGGCGGCGGTCCCAAT 774
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84

775 CGAGCGCGGCGGTTCGTTGCGAGCGGCGGTGCGGCGGCGGCGGCTC 824
|||||
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-639

seq_documentation_block:
; Sequence 639, Application US/10080170B
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170B
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 639
; LENGTH: 100
```

```
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170B-639

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 99.000

alignment_block:
US-09-462-480-1 x US-10-080-170B-639 ..
Align seg 1/1 to: US-10-080-170B-639 from: 1 to: 100

525 ATGCAGAGATGAAGACCGATGCGCTACCTCGGCGAGGAGCGGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17

575 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGGTGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34

625 CGAGCGCGGTTCGTTGCGAGCGGCGGTGCGGCGGCGGCGGCGGCC 674
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50

675 GCCAGCGCGGTTCGTTGCGAGCGGCGGTGCGGCGGCGGCGGCGGCC 724
|||||
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67

725 GGAACCTCGACGATCTCGACGATATTCGTCAGGCGGCGGTCCCAAT 774
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84

775 CGAGCGCGGCGGTTCGTTGCGAGCGGCGGTGCGGCGGCGGCGGCTC 824
|||||
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-6

seq_documentation_block:
; Sequence 6, Application US/09116492A
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, A TOOL FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492A
; CURRENT FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-116-492A-6

alignment_scores:
  Quality: 247.00      Length: 49
  Ratio: 5.041        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-462-480-1 x US-09-116-492A-6 ..
```

Align seg 1/1 to: US-09-116-492A-6 from: 1 to: 49

```
525 ATGCAGAGATGAACACCGATGCGCTACCTCGCGGAGGAGGAGGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlnGluAlaGlyAs 17
|||||
575 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
|||||
625 CGACGGCAGGTCTGTCGAGGGCCAGTGGCGCGCGCGGGGAGC 671
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 49
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-6

```
seq_documentation_block:
; Sequence 6, Application US/10140045
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS METHODS THEREOF CAPABLE OF MODULATING VARIOUS RESPONSES
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/10/140,045
; PRIOR APPLICATION NUMBER: 2002-05-08
; PRIOR FILING DATE: 1998-07-16
; PRIOR FILING DATE: 1998-07-16
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-140-045-6
```

alignment_scores:
Quality: 247.00 Length: 49
Ratio: 5.041 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-480-1 x US-10-140-045-6 ..

Align seg 1/1 to: US-10-140-045-6 from: 1 to: 49

```
525 ATGCAGAGATGAACACCGATGCGCTACCTCGCGGAGGAGGAGGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlnGluAlaGlyAs 17
|||||
575 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
|||||
625 CGACGGCAGGTCTGTCGAGGGCCAGTGGCGCGCGCGGGGAGC 671
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 49
|||||
```

seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17077

```
seq_documentation_block:
; Sequence 17077, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE OF MODULATING VARIOUS RESPONSES
; TITLE OF INVENTION: MODULATING VARIOUS RESPONSES
; FILE REFERENCE: 2750-1481P
```

```
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 17077
; LENGTH: 827
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1...827
; OTHER INFORMATION: Ceres Seq. ID no. 2708208
US-09-935-625-17077
```

alignment_scores:
Quality: 210.50 Length: 380
Ratio: 1.126 Gaps: 23
Percent Similarity: 49.211 Percent Identity: 31.316

alignment_block:
US-09-462-480-1/rev x US-09-935-625-17077 ..

Align seg 1/1 to: US-09-935-625-17077 from: 1 to: 827

```
1253 TCTCAGAGTGGCTCAACAGCTATAACACGAGAAAGGCGAGACGCGG 1204
|||||
234 SerArgSerProIleArgArg.....HisArgArgProThrHisGluG 248
|||||
1203 AAGTCGAACCTCCCGATCCCGTGTTCGGCTATTCTACGCGAACTCGGC 1154
|||||
248 yArgArgGlnSerProAlaPro.....SerArgArgArgA 260
|||||
1153 GTTCGCCCTATGCGAACATCCCGAGTGAGTTCCTTCGGTCCGAGCCATTG 1104
|||||
260 rgSerProSerProProAlaArgArg.....ArgSerProSerPro 274
|||||
1103 CCTGACCGGCTTCGCTGATCGTCGCGCGAGTTCGCGCGCGTGTGTC 1054
|||||
275 ProAlaArgArgArgArgSerProSerProProAlaArgArgHisArgSe 291
|||||
1053 AGCTCGGTAGCGTGGCGTCCCTTTTGTGGACACCTCGGTAGCGCTC 1004
|||||
291 rProThrProProAlaArgGln.....ArgArgSerProSerProProA 306
|||||
1003 CGAACCGCTACCGC...CCGAGCGCGCTCGAGCTTGGTCAGGAGTCT 957
|||||
306 laArgArgHisArgSerProProAlaArgArgArgArgSerProSer 322
|||||
956 TCCCTCGTCAAGGAGGGAATGAATGGACGTGACATTTCCCTGGATTGCG 907
|||||
323 ProProAlaArgArgArgSerProSerProProAlaArgArgArgAr 339
|||||
906 CTTGCGCGGCGCTCGATACCGCGGAAATTCACCTGCTGCTGCTCATGTT 857
|||||
339 gSerProSerProLeuTyArgArgAsnArgSerPro..SerProLeuTy 355
|||||
856 TTTCGCTCGTTTCTTTTCGTATTAGCGGTCCAGAACCCATTTCGCGAGA 807
|||||
355 rArgArgAsn.....ArgSerArgSerProLeuAlaIlysa 367
|||||
806 CAGCGCTGCTGCTGCTCTCT...CGTCGCGCTCGAGTATTGGACGCGCG 760
|||||
367 rgGlyArgSerAspSerProGlyArgSerProSerProValAlaArgLeu 383
|||||
759 CCTGACGAAATATTGTCGAGA.....TCTGCTCGAGT 728
|||||
384 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 400
|||||
727 TCCTGCTCTCTCTTATTGGCTGCTTCTTGAAGCGCACCG...CGGC 681
|||||
400 uprSerProProValAlaGlnArgLeuProSerProProProArgArgA 417
|||||
680 CTGGCGGCGCGCTCCCGCGCGCGCGCTGCGCTGCAACGAACCTG 631
|||||
```



```
513 ....GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 528
432 ..TCACAGTCGCTCTTCGTCCTCCAGTCGTCCTCGCTTCCTTTCACG 385
||||| ||| : : : : : ||||| |||
528 etSerProVal.ArgGlyArgGlyLysSerProSerSerArg.HisG 544
||||| : : : : : ||| : : : : :
384 CTCCTCGCGAGCGGTGCCGCGCAGCACACCGG.....GGCTGG 344
||||| : : : : : ||| : : : : :
544 lnLysAlaArgSerProValArgArgSerProThrProValAsnArg 560
343 TGGAGCGCGCGGATGGACCCCTGGCCCATCGTCCCGGACCCACCGGA 294
||||| : : : : : ||| ||||| : :
561 ArgSerArgSerSerSerAlaSerArgSer....ProAspArgArgAr 576
293 GCGGCGCCACCGG.....TCACGAGCATCCGCGCAACAGCCGC 256
||||| ||| ||||| : : : : : : : : : : :
576 gArgArgSerProSerSerSerArgSerProSerArgSerArgSerProp 593
255 CGGCA.....TCACCGAGGGGGCAA 236
||||| ||||| ||||| : : : : :
593 roValLeuHisArgSerProSerProArgGlyArg 604

seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17480

seq_documentation_block:
; Sequence 17480, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; OF INVENTION: MODULATING VARIOUS RESPONSES
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 17480
; LENGTH: 859
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: peptide
; LOCATION: 1..859
; OTHER INFORMATION: Ceres Seq. ID no. 3023746
US-09-935-625-17480
```

```
alignment_scores:
  Quality: 210.50      Length: 380
  Ratio: 1.126         Gaps: 23
  Percent Similarity: 49.211      Percent Identity: 31.316

alignment_block:
US-09-462-480-1/rev x US-09-935-625-17480 ..
Align seg 1/1 to: US-09-935-625-17480 from: 1 to: 859

1253 TCTCAGATCGCTCAACGCTATAACACGAGAAAGGCGGAGACCGCG 1204
||||| ||| : : : : : ||||| |||
266 SerArgSerProLleArgArg.....HisArgArgProThrHisGluG 280
1203 AAGTCAACTCGCCGATCCGCTGTTTCGCTATCTACGGAACTCGGC 1154
|||| : : : ||||| |||
280 yArgGlnSerProAlaPro.....SerArgArgArg 292
1153 GTTCCCTATCGGAACATCCAGTGAGCTTCCTTCGTCGCAACCCATTG 1104
|||| ||| : : : : : ||| |||||
292 rgSerProSerProAlaArgArg.....ArgSerProSerPro 306
1103 CCTGACCGGCTTCGCTGATCGTCGCGCCAGGTTCTCAGCGCGTGTTC 1054
|||| ||| ||| : : : : : ||| : :
307 ProAlaArgArgArgSerProSerProProAlaArgHisArgSe 323
1053 AGTCGGTAGCGGTGGCTCCATTTTGTGTGGACACCCCTGGTACGGCTC 1004
||| ||| |||
```

```
323 rProThrProProAlaArgGln.....ArgArgSerProSerProProA 338
1003 CGAACCCCTACCGC...CCAGGCGCGCTGGAGCTTGGTCAGGACTGCT 957
||||| ||| ||| ||| |||
338 laArgArgHisArgSerProProAlaArgArgArgArgSerProSer 354
956 TCCCTCGTCAAGGAGGAATGAATGACGCTGACATTTCCCTGGATTGCG 907
||| : : : : : : : : : : :
355 ProProAlaArgArgArgArgSerProSerProProAlaArgArgArgAr 371
906 CTTCGCCGCGCGCTGATACCCGCGGAAATTCACCTGCTGCTGCTCATGTT 857
||| ||| ||| ||| ||||| : : ||| : : : : :
371 gSerProSerProLeuTyArgArgAsnArgSerPro..SerProLeuTy 387
856 TTTGCTCCGTTTCTTTTCGTATTAGCGGTGTCAGAGCCCATTTTCGGAGA 807
: : : ||| ||||| ||||| : : : : :
387 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 399
806 CAGCGCTGCTGCTGCTCT...CGTCGCGCCTCGAGTATTGGACGCGG 760
: : : : : ||| ||||| : : : : :
399 rgGlyArgSerAspSerProGlyArgSerProSerProValAlaArgLeu 415
759 CTTGACCAATATTCGTCGAGA.....TCTCGTCGAGT 728
||| : : : : : |||
416 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 432
727 TCTGCTTCTGCTTATTGGTCTTCTTGGAGCGCACCCACG...CGGC 681
: ||| : : ||| ||| ||| ||||| |||
432 uProSerProProValAlaGlnArgLeuProSerProProArgArgA 449
680 CTGGCGCGCGCTCCCGCGCGCGCCACTGGCCCTGCAACAACTG 631
||| ||||| |||||
449 laGlyLeuProSerProPro.....ProAlaGlnArgLeu 460
630 CCGTCGACTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 586
||||| ||||| ||||| : : : : :
461 ProSerProProProArgArgAlaGlyLeuProSerProMetArgIleG 477
585 .....TCGCTCGAAATACCTG..... 568
||| ||||| |||||
477 yGlySerHisAlaAlaAsnHisLeuGluSerProSerProSerSerLeuS 494
567 ..CCTCTGCGCGAGGTAGCGCATCGTCTTCATCTGCCATGCTGG 520
||||| : : : : : ||| |||
494 erProGlyArg.....LysLysValLeuProSerProProValArg 508
519 ACTACTTCTCTCTTACCTTCCTCGCCAAATGTTGCCAAGTCTTCCGG 470
||||| : : : : : ||||| : : : : :
509 ArgArgArgSerLeuThrProAspGluGluArgValSerLeuSerGln.. 524
469 CCGGTTGGCGGGAAGTCTGTTGTCATTAC.....GGGAGC. 433
||||| : : : : : ||| |||||
525 ....GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 540
432 ..TCACAGTCGCTCTTCGTCCTCCAGTCGCTGCTGCTGCTTCTTACG 385
||||| : : : : : ||||| |||||
540 etSerProVal.ArgGlyArgGlyLysSerProSerSerArg.HisG 556
384 CTCCTCGCGAGCGGTGCCGCGGACGACACCGG.....GGCTGG 344
||||| : : : : : ||| : : : : :
556 lnLysAlaArgSerProValArgArgArgSerProThrProValAsnArg 572
343 TGGAGCGCGCGGATTCGGAACCTTCGCTGCGGACCCACCGCA 294
||||| ||| : : : : : ||| |||||
573 ArgSerArgArgSerSerAlaSerArgSer...ProAspArgArgAr 588
293 GCGGCGCCACCGG.....TCACGAGCATCCGCGCAACAGCGCG 256
||||| ||| ||||| : : : : :
588 gArgArgSerProSerSerArgSerProSerArgSerArgSerProp 605
255 CGGCA.....TCACGAGGGGGCAA 236
||||| ||||| ||||| : : : : :
605 roValLeuHisArgSerProSerProArgGlyArg 616
```


US-09-935-625-17075

alignment_scores:

Quality: 210.50 Length: 380
Ratio: 1.126 Gaps: 23
Percent Similarity: 49.211 Percent Identity: 31.316

alignment_block:

US-09-462-480-1/rev x US-09-935-625-17075

Align seg 1/1 to: US-09-935-625-17075 from: 1 to: 891

```
1253 TCTCAGAGTGGCTCAACAGTATAAACACAGAGAAAGCGGAGACGCGG 1204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 SerArgSerProIleArg,.....HisArgArgProThrHisGluG1 312
1203 AAGTCGAACTCGCGGATCCGCTGTTGCTATTCTACGGAACCTGGC 1154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 yArgArgGlnSerProAlaPro.....SerArgArgArgA 324
1153 GTTGCCTATCGAATCCAGTGACGTTGCTTCGGTCCGAAGCCATTG 1104
|||:|||||:|||||:|||||:|||||:|||||:|||||:
324 ySerProSerProAlaArgArg.....ArgSerProSerPro 338
1103 CTGACCGGCTTCGCTGATCGTCGCGGCGGAGTTCTGCAGCGCTGTTC 1054
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 ProAlaArgArgArgSerProSerProProAlaArgArgHisArgSe 355
1053 AGTCGGTAGCGGTGGCGTCCCTTTTGTGGACACCTCGTACGCGCTC 1004
|||:|||||:|||||:|||||:|||||:|||||:|||||:
355 rProThrProProAlaArgGln.....ArgArgSerProSerPro 370
1003 CGAACCGCTACCGC...CCAGCGCGTCCGAGCTTGCTCAGGAGCTCT 957
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 laArgArgHisArgSerProProAlaArgArgArgArgSerProSer 386
956 TCCCTCTCAAGGAGGAATGAATGGACGTGCACATTTCCCTGGATTGCG 907
|||:|||||:|||||:|||||:|||||:|||||:|||||:
387 ProProAlaArgArgArgSerProSerProProAlaArgArgArgAr 403
906 CTGCGCGGGCTCGATACCGCGGAATTCACCTGCTCTGTGATGTTT 857
|||:|||||:|||||:|||||:|||||:|||||:|||||:
403 gSerProSerProLeuTyArgArgAsnArgSerPro..SerProLeuTy 419
856 TTGCTCGCTTCTTTTGTATTAGCGGTGACGAAGCCATTTCGCGAGA 807
|||:|||||:|||||:|||||:|||||:|||||:|||||:
419 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 431
806 CAGCGCTGCTGCTCTCT...CGTCGGCTCGAGTATTGGACGCGG 760
|||:|||||:|||||:|||||:|||||:|||||:|||||:
431 rgGlyArgSerAspSerProGlyArgSerProSerProValaArgLeu 447
759 CTTGACGAATATTCGTCAGA.....TCTCGTCGAGT 728
|||:|||||:|||||:|||||:|||||:|||||:|||||:
448 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 464
727 TCTGCTTCTGCTATTGCTGCTCTTGTGAAGCGCACCCAG...CGGC 681
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 uProSerProProValAlaGlnArgLeuProSerProProArgArgA 481
680 CTGGCGCGCTCCCGCGCGCGCGCCACTGGCGCTGCAACGAACCTG 631
|||:|||||:|||||:|||||:|||||:|||||:|||||:
481 laGlyLeuProSerProPro.....ProAlaGlnArgLeu 492
630 CGGTGCTCCACCTGTCGATGCTGGTTTTCAGTCCGCGGAGA..... 586
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 ProSerProProProArgArgAlaGlyLeuProSerProMetArgIleG1 509
585 .....TCCGCTCGAATATTACCTG..... 568
509 yGlySerHisAlaAlaAsnHisLeuGluSerProSerProSerLeuS 526
```

```
567 ..CCTCTGCCCCGAGGTAGCGCATCGGTCTTCATCTCTGCCATCGTGG 520
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
526 erProProGlyArg,.....LysLysValLeuProSerProProValArg 540
519 ACTACTTTCTCTTACCTTCCTCCGCCAAATGTTGGCAAGTCTTCGGG 470
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 ArgArgArgSerLeuThrProAspGluGluArgValSerLeuSerGln.. 556
469 CCCGGTGGCGGGAAGTCTGTTGCTATTAC.....GGGAGC. 433
|||||:|||||:|||||:|||||:|||||:|||||:
557 ....GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 572
432 ..TCACCAAGTCGCTCTTCCTCCAGTCGCTCGTCTGCTCTTCTTCACG 385
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
572 etSerProVal.ArgGlyArgGlyLysSerProSerSerArg.HisG 588
384 CTCCTGCGGAGCGGTGCGCGCGGACGACCGC.....GGCTGG 344
|||||:|||||:|||||:|||||:|||||:|||||:
588 lnLysAlaArgSerProValArgArgArgSerProThrProValAsnArg 604
343 TGGAGCGCGCGGATTGGCAACCTCGCTCCGCGGACCCACCGCA 294
|||||:|||||:|||||:|||||:|||||:|||||:
605 ArgSerArgArgSerSerSerAlaSerArgSer...ProAspArgArgAr 620
293 GCGCGCGCACCGC.....TCACCAAGTCGCGGCAACGACGCGC 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
620 gArgArgSerProSerSerArgSerProSerArgSerArgSerProp 637
255 CGGCA.....TCACCGAGGGGCGCA 236
|||||:|||||:|||||:|||||:|||||:|||||:
637 roValLeuHisArgSerProSerProArgGlyArg 648
```

seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17478

seq_documentation_block:

Sequence 17478, Application US/09935625

GENERAL INFORMATION:

APPLICANT: N. ALEXANDROV et al.

TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAP.

TITLE OF INVENTION: MODULATING VARIOUS RESPONSES

FILE REFERENCE: 2750-1481P

CURRENT APPLICATION NUMBER: US/09/935,625

CURRENT FILING DATE: 2001-08-24

NUMBER OF SEQ ID NOS: 33136

SEQ ID NO 17478

LENGTH: 903

TYPE: PRT

ORGANISM: Arabidopsis thaliana

FEATURE:

NAME/KEY: peptide

LOCATION: 1..903

OTHER INFORMATION: Ceres Seq. ID no. 3023744

US-09-935-625-17478

alignment_scores:

Quality: 210.50 Length: 380
Ratio: 1.126 Gaps: 23
Percent Similarity: 49.211 Percent Identity: 31.316

alignment_block:

US-09-462-480-1/rev x US-09-935-625-17478

Align seg 1/1 to: US-09-935-625-17478 from: 1 to: 903

```
1253 TCTCAGAGTGGCTCAACAGTATAAACACAGAGAAAGCGGAGACGCGG 1204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 SerArgSerProIleArg,.....HisArgArgProThrHisGluG1 324
1203 AAGTCGAACTCGCGGATCCGCTGTTGCTATTCTACGGAACCTGGC 1154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 yArgArgGlnSerProAlaPro.....SerArgArgArgA 336
```

```
1153 GTTGCCCTATGCGAACAATCCAGTACGTTGCTTTCGGTCCGAAGCCATTG 1104
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 rgSerProSerProProAlaArgArg.....ArgSerProSerPro 350
1103 CTGACCGCTTCGTCGTATGTCGCCGCCAGGTCTCTGCAGCCGCTGTTTC 1054
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 ProAlaArgArgArgSerProProAlaArgArgHisArgSe 367
1053 AGCTCGGTAGCGTGGCTGCCATTTTGTCTGGACACCTGTCAGCCCTC 1004
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 rProThrProProAlaArgGln.....ArgArgSerProSerProProA 382
1003 GCAACCGTACGCG...CCAGCGCGCTGCGAGCTTGTTCAGGACTGCT 957
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 laArgArgHisArgSerProProAlaArgArgArgSerProSer 398
956 TCCCTCTGTCAGGAGGAAGTAATGGAGCTGCACATTTCCCTGGATTGCG 907
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399 ProProAlaArgArgArgArgSerProSerProProAlaArgArgArg 415
906 CTTCGCGCGGCTCGATACCCGCGAAATTCACCTGCTGCTGTCATGTT 857
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
415 gSerProSerProLeuTyArgArgAsnArgSerPro...SerProLeuTy 431
856 TTTCGCTCGTTCTTTCGTATTACGGGTGAGAGCCCATTTGCGAGGA 807
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 443
806 CAGCGCTGCTGCTGCTCT...CGTCGCGCTCGAGTATTGGAGCCGG 760
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
443 rgGlyArgSerAspSerProGlyArgSerProSerProValAlaArgLeu 459
759 CTTGACGAATATTCTCGAGA.....TCTCGTCGAGT 728
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
460 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 476
727 TCCTGCTTCTGCTATTGCTGCTTCTTTGGAAGCGCACCCG...CGCG 681
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
476 uProSerProProValAlaGlnArgLeuLeuProSerProProArgArg 493
680 CTGGCGCGCTGCCCGCGCGCGCCACTGGCCCTGCAACGACCTG 631
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
493 laGlyLeuProSerProPro.....ProAlaGlnArgLeu 504
630 CCGTCACTCACCTGCTGATCTGGTTTCAGGTCGCCCGGAGA..... 586
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
505 ProSerProProProArgArgAlaGlyLeuProSerProMetArgIleG 521
585 .....TCCGCTCGAATTACCTG..... 568
521 yGlySerHisAlaAlaAsnHisLeuGluSerProSerProSerLeuS 538
567 ..CTTCCTGCCGAGGGTAGCGCATCGTCTTCATCTCTGCGATGCTGG 520
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
538 erProProGlyArg.....LysLysValLeuProSerProProValArg 552
519 ACTACTTCTCTTCTTCTCTTCTCTGCGCAAAATGTTGCAAGTCTTCGG 470
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
553 ArgArgArgSerLeuThrProAspGluGluArgValSerLeuSerGln.. 568
469 CCCGGTGGCGGAGTCTCTTGTTCATTAC.....GGGAGC. 433
569 ....GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 584
432 ..TCACCACTGCTCTCTCTCTGCTCCAGTCGCTGCTCTCTTCTTCCAG 385
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
584 etSerProVal..ArgGlyArgGlyLysSerSerProSerArg.HisG 600
384 CTCTGTCGCGAGCGTGCAGCGCGGACAGACCCG.....GGCTGG 344
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
600 lnLysAlaArgSerProValArgArgArgSerProThrProValAsnArg 616
343 TGGAGCCCGCGATTGCGAACCTTGGCCCATCTGCTCCGGACCCACCGGA 294
```

```
617 ArgSerArgArgSerSerAlaSerArgSer...ProAspArgArgAr 632
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
293 GCGCGCCACCCG.....TCACCGACGATCCGGCAACAGCCGC 256
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
632 gArgArgSerProSerSerArgSerProSerArgSerArgSerProp 649
255 CGGCA.....TCACCGAGGGGCAA 236
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
649 roValLeuHisArgSerProSerProArgGlyArg 660
seq_name: /cgn2_5/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-7
seq_documentation_block:
; Sequence 7, Application US/09116492A
; GENERAL INFORMATION:
; APPLICANT: GICOUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MY
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, A
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492A
; PRIOR FILING DATE: 1998-07-16
; PRIOR FILING DATE: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-116-492A-7
alignment_scores:
Quality: 205.00 Length: 42
Ratio: 4.881 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-462-480-1 x US-09-116-492A-7 ..
Align seg 1/1 to: US-09-116-492A-7 from: 1 to: 42
699 CAAGNAGCAGCCAATAAGCAGAAGCAGAACTCGACGAGATCTCGACGAA 748
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1 GlnGluAlaAlaAsnLysGlnLysGlnLeuAspGluIleSerThrAs 17
749 TATTGCTCAGCGCGCTCCCAATCTCGAGGCGCCGACGAGCAGCAGCAGC 798
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
17 nIleArgGlnAlaGlyValGlnTySerArgAlaAspGluGlnGlnG 34
799 AGGCGCTGCTCTCGCAATGGGCTTC 824
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
34 lnAlaLeuSerSerGlnMetGlyPhe 42
seq_name: /cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-7
seq_documentation_block:
; Sequence 7, Application US/10140045
; GENERAL INFORMATION:
; APPLICANT: GICOUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MY
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, A
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/10/140,045
```



```
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: US/09/116.492A
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-140-045-7
```

```
alignment_scores:
  Quality: 205.00      Length: 42
  Ratio: 4.881         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-462-480-1 x US-10-140-045-7
Align seg 1/1 to: US-10-140-045-7 from: 1 to: 42
```

```
699 CAGAGAGCACCATAAGCAGAGCAGGAACTCGACGAGATCTCGACGAA 748
|||||
1 GlnGluAlaAlaAsnLysGlnLysGlnGluLeuAspGluLeuSerThrAs 17
749 TATTCGTACAGCGCGCGTCCAAATCTCGAGGGCGCGAGGACGACGAGC 798
|||||
17 nileArgGlnAlaGlyValGlnTyrSerArgAlaAspGluGlnGlnG 34
799 AGCGCGTGTCTCGCAATGGGCTTC 824
|||||
34 lnAlaLeuSerSerGlnMetGlyPhe 42
```

seq_name: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2231

```
seq_documentation_block:
; Sequence 2231, Application US/60389987
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 66088.465p2
; CURRENT APPLICATION NUMBER: US/60/389,987
; NUMBER OF SEQ ID NOS: 3025
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2231
; LENGTH: 2263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-389-987-2231
```

```
alignment_scores:
  Quality: 203.50      Length: 442
  Ratio: 0.988         Gaps: 27
Percent Similarity: 46.606 Percent Identity: 28.733
```

```
alignment_block:
US-09-462-480-1 x US-60-389-987-2231
Align seg 1/1 to: US-60-389-987-2231 from: 1 to: 2263
```

33 GGGCGGACCGGGCGGCGCAACCCAGCGACGAGGA...AGCGCGCGAGA 79

```
|||||
448 GlySerHisArgGluLeuSerSerProThrSerLysAsnArgSerH1 464
80 TGGCGCTGCTCGGCACCGCTGTCGAACCATCGCTGGCTGGTGA 129
|||||
464 sGlyArgAlaLysArgAspLysSer.HisSerHisThr..... 476
130 TCAGGCCCCAGCGCGCGCGCTGCTCGCGCGAGTCTGCTACCTGG 179
|||||
477 .....ProSerArgMetGlyArgSerArgSerProAlaThrAlaL 491
180 CGCAGGTGGTGTGACCGCGCGCGGTGATGTCACGTATCGAAA 229
|||||
491 sArgGlyArgSerArgSerArgThrProThrLysArg.GlyHisSerArg 507
230 AGCGGTTGCCCTCGTGTGATCGCGCGCTGTTGCCGGATCGTCGGT 279
|||||
508 SerArgSerPro..... 511
280 ACGGCTGGCGCGCTCGGTGGTCCGG.....GAGCGATGGG 317
|||||
512 .....GlnTrpArgArgSerArgSerAlaGlnArgTrpG 523
318 CCAGGTTCCGAATCCG.....GCGGCTCCACGACC 349
|||||
523 lyArgSerArgSerProGlnArgGlyArgSerArgSerProGlnArg 539
350 CGGCTGTGTCGCGCGCGCTCGCGAGCGTGAAGAGACGAC 399
|||||
540 ProGlyTrpSer.....ArgSerArgAsnThrGln.ArgArgGlyA 553
400 GAGGAGACTGGGACGAGGACGACTGCTGAGTCCCGTAATGACAAC 449
|||||
553 rgSerArgSerAlaArgArgGly.....ArgSerHisSer 564
450 AGACTTCCCGCACCGCGCGGAGAACTTGCCAAACATTTTGGCGGAG 499
|||||
565 Arg.SerProAlaThrArg..GlyArg..... 572
500 AGGTAAGAGAGAAAGTAGTCCAGCATGGCAGAGATCAAGACCGATCCG 549
|||||
573 .....SerArgSerArgThrProAla..... 579
550 CTACCTCGGCGAGGAGGAGGTAATTTTCGAGCGATCTCGGCGACCTG 599
|||||
580 .....ArgArgGlyArgSerArgSerArgThrProAlaArgAr 592
600 AAAACCCAGATCGACAGGTGGAGTCGACGCGAGTTCTGTCAGGCCA 649
|||||
592 gArgSerArgSerArgThrProThrArgArgSerArgSerArgThr. 608
650 GTGGCGGCGCGCGGAGCGCGCGCGCGCGCGCGCGTGGTGGCTTCC 699
|||||
609 .....ProAlaArgArgGlyArg.....Ser 615
700 AAGAAGCAGCCAATAAGCAGAAGCAGGAACTCGACGAGATCTCGACGAAT 749
|||||
616 ArgSerArgThrProAlaArgArgSerArgSerArgThrArgSerPro..... 630
750 ATTCTGTCAGCGCGGTCCCAATCTCGAGGGCGCGAGGAGGAGCAGCA 799
|||||
631 ....ValArgArgSerArgSerArgSerProAlaArgSerGlyA 646
800 GCGCTGTCTCTCGAAATGGGCTTCTGACCCCGCTAATACGAAAAGAACG 849
|||||
646 rgSer.....ArgSerArgThr 651
850 GAGCAAAACATGACAGCAGCAGCAGTGAATTTTCGGCGGTATCGAGCGCG 899
|||||
652 ProAlaArgArgGlyArgSerArgSerArgThrProAlaArgArgGlyAr 668
900 CGGCAAG...CGCAATCCAGGGAAATGTCACGTCCTATTCATTCCTCCTT 946
|||||
```

[illegible]

```

seq_documentation_block:
; Sequence 13, Application US/10080170B
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TREATMENT OF MYCOBACTERIOSES
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170B
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 302
; TYPE: PRS
; ORGANISM: Mycobacterium leprae
US-10-080-170B-13

```

```

alignment_scores:
  Quality: 198.00      Length: 126
  Ratio: 2.329         Gaps: 4
  Percent Similarity: 67.460      Percent Identity: 41.270

alignment_block:
US-09-462-480-1 x US-10-080-170B-13 ..

Align seg 1/1 to: US-10-080-170B-13 from: 1 to: 302

1 CTCACAGAGGTGACGTGGTTTTCAGCCAGGTGGCGGCACCGCGCGCG 50
   :||||||| :||||||| :|||: :|||:
190 MetClnValLysSerLeuPheThrSerIleAspSerThrGlyValty 206
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

51 CACCCAGCGGAC.....GAGGAAGCGCGCAGATGGGCGCTGC 88
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

```

OM of: US-09-462-480-1 to: PIR_71:* out_format : pfs
Date: Jul 22, 2002 1:25 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=framet_n2p.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09462480/runat_18072002_164418_19445/app_query.fasta_1.2850
-DB=PIR_71 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -XGAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEA=SIZE=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09462480.@CGN1_1_73
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-462-480-1
Query length: 1277
Database: PIR_71:*
Database sequences: 283138
Database length: 96089334
Search time (sec): 87.620000

score_list:	Strd	Orig	ZScore	Escore	Len	Documentation
PIR2:G70802	+	713.00	852.17	6.1e-40	368	! Probable PPE protein - Mycobact
PIR2:H70802	+	486.00	591.14	7.7e-25	100	! hypothetical protein Rv3874 - M
PIR2:A70803	+	479.50	583.18	2.3e-24	95	! early secretory antigen target e
PIR1:Q0863	+	212.50	248.07	1.5e-06	660	! BHLF1 protein - human herpesvir
PIR2:G84693	+	210.50	243.17	2.1e-06	391	! probable proline-rich protein
PIR2:T10033	+	198.00	237.17	1.3e-05	302	! hypothetical protein MLCB628.14
PIR2:T02345	+	195.50	219.46	2.2e-05	1791	! hypothetical protein KIAA0324
PIR2:T10032	+	187.00	232.36	6.8e-05	100	! hypothetical protein MLCB628.13
PIR2:T10031	+	186.00	232.36	7.9e-05	95	! early secretory antigen target e
PIR2:T13078	+	185.00	213.84	0.0001	772	! KIAA0992 protein - human (fragm
PIR2:S52796	-	182.50	216.22	0.0001	403	! prpL2 protein - human (fragm
PIR2:T02345	+	179.50	200.30	0.0003	1791	! hypothetical protein KIAA0324
PIR2:B40505	+	178.50	198.37	0.0003	1958	! hypothetical protein - suid he
PIR2:T08179	+	176.00	204.62	0.0004	640	! LRG5 protein - chlamydomonas re
PIR2:A45748	+	174.00	199.22	0.0006	920	! collagen alpha 1(VII) chain - m
PIR2:J00405	+	172.50	195.91	0.0007	1106	! hypothetical 119.5K protein (u
PIR1:B45344	+	172.50	192.19	0.0007	1733	! probable nuclear antigen - su
PIR2:S27923	+	171.00	195.60	0.0009	924	! gene Lr3 protein - human herpes
PIR2:S28774	+	171.00	194.72	0.0009	1027	! collagen alpha chain - tube wc
PIR1:EBE1F	+	170.50	191.22	0.0010	1460	! immediate-early protein IE180
PIR2:F75311	+	167.00	195.06	0.0016	552	! ABC transporter, ATP-binding pr
PIR2:S0832	-	165.00	186.36	0.0023	1184	! atrophin-1 - human
PIR2:G01763	-	164.50	185.76	0.0025	1184	! atrophin-1 - human
PIR2:T31611	+	164.50	183.35	0.0025	1585	! hypothetical protein Y508A.g
PIR2:A54849	+	164.00	177.64	0.0028	2944	! collagen alpha 1(VII) chain pr
PIR2:A59266	+	163.50	175.54	0.0031	3530	! unconventional myosin-15 - hum
PIR2:T01696	-	162.50	191.82	0.0032	426	! pistil extensin-like protein ph
PIR2:J04859	-	162.50	186.22	0.0033	839	! extensin homolog F28A21.80 - A
PIR2:S51342	+	161.00	184.64	0.0042	817	! verprolin - yeast (Saccharomyce
PIR2:IA48103	+	160.00	178.15	0.0050	1549	! type VII collagen - Chinese ha
PIR1:A43291	+	159.50	178.55	0.0054	1373	! collagen alpha 2(I) chain pred
PIR2:G84348	+	159.00	186.40	0.0055	494	! hypothetical protein Vng1983h
PIR2:T43556	-	158.50	184.56	0.0060	574	! Wiskott-Aldrich syndrome prote
PIR2:T38819	-	158.50	184.56	0.0060	574	! Wiskott-Aldrich syndrome prote
PIR1:A45344	-	158.50	176.93	0.0063	1446	! immediate-early protein - suid
PIR1:CGHU2V	+	158.50	176.65	0.0063	1496	! collagen alpha 2(V) chain pred
PIR2:T45134	+	157.50	184.04	0.0069	529	! hypothetical protein [imported]
PIR1:EBE1D	+	157.50	180.37	0.0071	825	! immediate-early protein RL2 - H
PIR2:A36068	+	157.00	180.92	0.0076	718	! major ampullate fibroin protein
PIR2:T05352	+	157.00	179.46	0.0077	857	! hypothetical protein F8B4.120 -

PIR2:G84693 + 157.00 179.13 0.0077 891 ! probable proline-rich protei
PIR1:A45344 + 156.50 174.53 0.0086 1416 ! immediate-early protein - s
PIR2:T46289 + 156.00 178.21 0.0090 862 ! hypothetical protein DKF2P4
PIR1:CGCH2S + 156.00 177.29 0.0091 964 ! collagen alpha 2(I) chain p
PIR1:CGCH1S + 156.00 176.64 0.0091 1042 ! collagen alpha 1(I) chain
seq_name: pir2:G70802
seq_documentation_block:
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70802
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordo
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen
A:Reference number: A70500; MUID:98295987
A:Accession: G70802
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-368 <COL>
A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAA17965.1; PID:e12
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

alignment_scores:	Quality:	713.00	Length:	143
	Ratio:	5.057	Gaps:	0
Percent Similarity:	98.601	Percent Identity:	96.503	
alignment_block:				
US-09-462-480-1 x G70802	..			
Align seg 1/1 to: G70802 from: 1 to: 368				
1 CTGACGAGGTGACGCTGTTGTTTTCAGCCAGGTGGCGCGCACCGCGCGGG 50				
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG1 242				
51 CAACCCACGCGCACGAGGAGCGCGACATGGCGCTGCTGCGCACGACGTC 100				
242 YAsnProAlaaspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259				
101 CGCTGTCGAACATCCGCTGGCTGGTGGATCAGGCCCGCGCGCGCGCG 150				
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275				
151 GGCCTGCTGCGCGGAGTCCCTACCTGCGCGCAGGTGGTGGTTCACCCG 200				
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292				
201 CACGCGCTGATGTCCTACGCTGATCGAAAGCGGTTGCCCTCCCTCGGTGA 250				
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309				
251 TCCCGCGCGGCTGTTCCCGGATCGTGGTGGTGGCGGGTGGCGCGCTCCG 300				
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325				
301 GGTCCGGAGGAGTGGCGGAGGTTCGGAATCCGCGCGCTCCACGAGCCC 350				
326 GlyAlaGlyAlaMetGlyGlnGlnGlyAlaGlnSerGlyGlySerThrArgPr 342				
351 GGGTCTGCTGCGCGGCGCACGCTCGCGCAGGAGGTGAAGAACGACGACG 400				
342 oGlyLeuValAlaProAlaProLeuAlaGlnGlnuargGluGluaspPsg 359				
401 AGGACGACTGGGACCGAAGAGGACGACTGG 429				

|||||
359 luAspAspTrpAspGluAspAspTrp 368

seq_name: pir2:H70802

seq_documentation_block:
hypothetical protein Rv3874 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70802
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: H70802
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-100 <COL>
A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL17966.1; PID:g296022
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3874

alignment_scores:
Quality: 486.00 Length: 100
Ratio: 4.860 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:

US-09-462-480-1 x H70802 ..

Align seg 1/1 to: H70802 from: 1 to: 100

525 ATGGCAGAGATGAGACCGATCGCGCTACCTCGGCGAGGAGGCGAGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlys 17
575 TTTCGAGCGGATCCCGCGACCTCAAAACCCAGATCGACCACTGGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
625 CGAGCGGAGTTCGTTGAGCGCCAGTCGCGCGCGCGCGGCGGAGCGCC 674
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
675 GCCAGGCGCGGCGTGTTCGCGTCCAGAGACGAGCCCAATAAGCAGAGCA 724
|||||
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
725 GGAACCTCAGCAGATCGCAGCATATTCGTCAGCGCGGCGTCCAATACT 774
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
775 CGAGGCGCGAGCAGCAGCAGCGCGCTCTCCGCAATGGGCTC 824
|||||
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: pir2:A70803

seq_documentation_block:
early secretory antigen target esat6 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70803; S49174
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: A70803
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-95 <COL>
A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL17967.1; PID:e126
A:Experimental source: strain H37RV
R:Soerensen, A.L.; Nagai, S.; Houen, G.; Andersen, P.; Andersen, A.B.
submitted to the EMBL Data Library, June 1994
A:Reference number: S49174
A:Accession: S49174
A:Molecule type: DNA
A:Residues: 1-13,'R',15-22,'S',24-95 <SOE>
A:Cross-references: EMBL:X79562; NID:g531708
C:Genetics:
A:Gene: esat6

alignment_scores:

Quality: 479.00 Length: 95
Ratio: 5.042 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-462-480-1 x A70803 ..

Align seg 1/1 to: A70803 from: 1 to: 95

860 ATGACAGACGACGAGTGGAAATTCGCGGGTATCGAGCGCGGCGAAGCGC 909
|||||
1 MetThrGluGlnGlnTrpAsnPheAlaGlyIleGluAlaAlaAlaSerAl 17
910 AATCCAGGGAATGTCACGTCCATTTCCTCTTCGACGAGGGAAGC 959
|||||
17 alleGlnGlyAsnValThrSerIleHisSerLeuLeuAspGluGlyLysG 34
960 AGTCCCTGACCAAGCTCGACGCGCTCGGCGGTAGCGGTTTCGGAGCGC 1009
|||||
34 InSerLeuThrLysLeuAlaAlaAlaTrpGlyGlySerGlySerGluAla 50
1010 TACAGGCGTCCAGCAAAATGGACCGCGCTACCGAGCTGAACAA 1059
|||||
51 TyrGlnGlyValGlnGlnLysTrpAspAlaThrAlaThrGluLeuAsnAs 67
1060 CGCGTTCAGAACTCGCGCGACGATCAGCGAAGCGGTGAGCAATGG 1109
|||||
67 nAlaLeuGlnAsnLeuAlaArgThrIleSerGluAlaGlyGlnAlaMetA 84
1110 CTTCGACCGAAGCAACGTCACCTGGGATGTTTCGCA 1144
|||||
84 laSerThrGluGlyAsnValThrGlyMetPheAla 95

seq_name: pir1:QBEB3

seq_documentation_block:
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C:Species: human herpesvirus 4, Epstein-Barr virus
C>Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
C:Accession: A03742
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr vi
A:Reference number: A93065; MUID:85035713
A:Accession: A03742
A:Molecule type: DNA
A:Residues: 1-660 <BAN>
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.
Nature 310, 207-211, 1984
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A:Reference number: A03794; MUID:84270667
A:Contents: annotation; protein coding region
C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399

C:Superfamily: human herpesvirus 4 BHLF1 protein

alignment_scores:

Quality: 212.50 Length: 560
Ratio: 0.944 Gaps: 29
Percent Similarity: 40.179 Percent Identity: 25.000

alignment_block:

US-09-462-480-1/rev x Q0BE3

Align seg 1/1 to: Q0BE3 from: 1 to: 660

```
1202 AGGTGCAACTCGCCGATCCCGTGT...CGCTATTCTACGCGAAC 1159
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
136 ArgGlyProArgProHisProAlaPheGlnValGlnTrpSerAlaArgAs 152
1158 TCGCGGTGGCCCTATGGACATCCAGTACGCTTCCGTCGGAAGC 1109
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 nProGlyCysPro...ArgThrTrpArgArgSerGlyAlaGlnArg 168
1108 CATTCGCTGACCGCTTCGCTGATCGTCGCGCCAGGCTTCGACGCGT 1059
||| ||||| ||||| ||||| ||||| ||||| |||||
168 lyHisPro.....ProGlyAlaGlyGlnArg 177
1058 TGTTCAGCTCGTAGCGGTGCGCTCCCATTTTGTGGACACCGCTGTAC 1009
||||| ||||| ||||| ||||| ||||| ||||| |||||
178 ProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyTh 194
1008 GCCT.....CCGAACGCTTACC 992
||||| ||||| ||||| ||||| ||||| ||||| |||||
194 rProAlaAlaProGlyProGlyGlyAlaAlaValProSerGlyAlaT 211
991 GCCCGAGCGCTCGAGCTGTGTGTCAGGACTGCTCCCTCGTCACGGA 942
||||| ||||| ||||| ||||| ||||| ||||| |||||
211 hrProHisProGluArgGly...SerGlyProAlaAspProProAlaAla 226
941 GGAATGAATGGACGTGACATTTCCCTGGATTGCGCTTGCGC..... 900
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
227 Ala.....ArgLeuProProGluAr 233
899 ....CGCCCTCGATACCGCGGAATTCACCTGCTGCTGTCATGTTTT 854
||||| ||||| ||||| ||||| ||||| ||||| |||||
233 gGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysProA 250
853 GTCCTGTTCTTTTCGATTAGCGGTCAGAACCCATT. GCGAGGACA 805
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 laGlyProProProThrArgSerGlyAlaAlaGlnArgThrHisArg 266
804 GCGCTGCTGCTGCTCCTCGTCGG..... 781
||| ||||| ||||| ||||| ||||| ||||| |||||
267 ArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArgTh 283
780 .....CCCTCGAGTATG 768
||| ||||| ||||| ||||| ||||| ||||| |||||
283 rTrpArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaG 300
767 GACCGCGCTGAGATATTCGAGATCTCGTCGAGTTCCTGCTTCT 718
||||| ||||| ||||| ||||| ||||| ||||| |||||
300 lyGlnArgPro.....SerGlyProThrGlyGlyArgProAlaAla 313
717 GCTTATTGGCTGCTTTTGAAGCGCACCGCGGCTGGCGCGCTC 668
||||| ||||| ||||| ||||| ||||| ||||| |||||
314 Pro.....GlyAlaProGlyThrProAlaAlaProGI 324
667 CCGC.....CCGCGCC 657
||| ||||| ||||| ||||| ||||| ||||| |||||
324 yProGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluA 341
656 GCGCACTGGCCCT.....GCAAGCACTGGCCG..... 628
||||| ||||| ||||| ||||| ||||| ||||| |||||
341 rgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlu 357
```

```
628 ..... 628
358 ArgGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysPr 374
627 ....TCGACTCCACCTGGTGCATCTGGTGTTCAGTCCGCGG..... 589
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
374 oAlaGlyProProProThrArgSerGlyAlaAlaAlaGlnArgThrHisA 391
588 .....AGATCCGCTCGAAT..... 574
||||| ||||| ||||| ||||| ||||| ||||| |||||
391 rgArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArg 407
573 .....TACCTGCTCCTGCGCG 557
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
408 ThrTrpArgArgArgSerGlyAlaGlnArgGlyHisProProGlyAl 424
556 AGGTAGCGGATCGGTCTTCATCTCGCCATGCTGGACTACTTCTCTC 507
||| ||||| ||||| ||||| ||||| ||||| |||||
424 aGlyGlnArg.....ProSerGlyProThrGlyGlyA 435
506 TTTACCTTCCTCGCCAAATGTTGGCAAGTCTCCGCGCGGGT...GGC 460
||| ||||| ||||| ||||| ||||| ||||| |||||
435 rgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGly 451
459 CGGAAGCTGTGTTCATTACGGAGCT.....CACCAGTCGCTCTT 416
||| ||||| ||||| ||||| ||||| ||||| |||||
452 .GlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlys 468
415 GGTCCAGTCGCTCCTGCTCTTCTTCACGC..... 384
||||| ||||| ||||| ||||| ||||| ||||| |||||
468 erGlyProAlaAspProAlaAlaAlaArgLeuProProGluArgGln 484
383 .....TCCTGCGCGAGCGG 370
485 GluProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGl 501
369 TCGCGCGCGACACCGCGGTGGTG.....GAGC 338
||| ||||| ||||| ||||| ||||| ||||| |||||
501 yProProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgp 518
337 GCGCGGATGGAAACCTGGCCCATCTCCGCGACCCACCG..... 296
||||| ||||| ||||| ||||| ||||| ||||| |||||
518 roProGlyCysProArgSerAlaArgAsnProGly.CysProArgThrTr 534
295 .....GAGCGCGCCACCGCTCACCGACGATCCCGC 265
534 pArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyG 551
264 AACAGCGCGGATCATCCGAGGGGCAACGGCTTTTCGATCAGCTGAG 215
||||| ||||| ||||| ||||| ||||| ||||| |||||
551 lnArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaPro 567
214 ACATCAGCGGGTGC.....GGTCAACGACCCACCTGCGCCAGGT 174
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
568 GlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProSerGl 584
173 AGCACTCCG.....CGCGCAGCAGCGCG..... 149
||||| ||||| ||||| ||||| ||||| ||||| |||||
584 yAlaThrProHisProGluArgGlySerGlyProAlaAspProProAla 601
148 .....GCGCGCGCTGGGCGCTGATCCAC 125
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 laAlaArgLeuProProGluArgGlnGluProArg..... 612
124 CAGCCAGCGGATGTTGACAGCGGACTGGTGGCAGCAGGCCCA..... 80
||| ||||| ||||| ||||| ||||| ||||| |||||
613 LeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProPro 629
79 .....TCTGCGCGCTTCTCTCTCGG...CTGGGTT 52
||||| ||||| ||||| ||||| ||||| ||||| |||||
629 oThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyC 646
51 GCCCGCGCGGTGCGCCACCTGGC 26
```



```
alignment_scores:
  Quality: 198.00      Length: 126
  Ratio: 2.329        Gaps: 4
  Percent Similarity: 67.460  Percent Identity: 41.270

alignment_block:
US-09-462-480-1 x T10033 ..
Align seg 1/1 to: T10033 from: 1 to: 302

1 CTGCAGCAGTCGCTGTTTCAGCCAGTGGCGGCACCGCGCGG 50
  :::::::::::::::::::::
190 MetGlnGlnValysSerLeuPheThrSerIleAspSerThrGlyValty 206
  :::::::::::::::::::::
51 CAACCCAGCCGAC.....GAGGAAGCCGCGCAGATGGCCCTGC 88
  :::::::::::::::::::::
206 rThrSerAlaGlnArgGlyAspThrGluSerAlaHisArgIleGlyLeup 223
  :::::::::::::::::::::
89 TCGGCACACGTCGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCC 138
  :::::::::::::::::::::
223 heGlyAlaSerThrLeuSerSerHisProLeuValGlyIleThrGlyThr 239
  :::::::::::::::::::::
139 AGCGCGGCGCGGCGCTGCTGCGCGGAGTCGCTACCTGCGCGCAGGTGG 188
  :::::::::::::::::::::
240 ThrThrAspThrArgLeuLeuCysAlaGluSerLeuProSerAlaSerG1 256
  :::::::::::::::::::::
189 GTCGTCGTCACCGCAGCGCTGATGCT.....CAGCTGATCAAAAGC 232
  :::::::::::::::::::::
256 ySerLeuAlaTrpThrProLeuMetThrGlnPheGlnLeuIleAspLys 273
  :::::::::::::::::::::
233 CGGTGTCGCCCTCG.....GTGATCGCGCGGCTGTTGCC 267
  :::::::::::::::::::::
273 erIleAlaProGluProArgGlnArgValMetLeuProTrpAlaAla 289
  :::::::::::::::::::::
268 GGATCGTGTGGTACGGGTGGCGCGCTCCGCTGGGTCCGGGAGGATGG 317
  :::::::::::::::::::::
290 GlySerPro.....G1 293

318 CCAGGGTTCGCAATCCGCGCTCCACC 345
  :::::::::::::::::::::
293 yHisAsnAlaGlnAspGlyGlyThr 302

.seq_name: pir2:T02345

seq_documentation_block:
hypothetical protein KIAA0324 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C:Accession: T02345
R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A:Description: Sequencing of human chromosome 16p13.3.
A:Reference number: Z14664
A:Accession: T02345
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1791 <RIC>
A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AA08453.1; PID:g2996650
C:Genetics:
A:Map position: 16
A:Introns: 1610/2; 1706/2
A>Note: KIAA0324

alignment_scores:
  Quality: 195.50      Length: 474
  Ratio: 0.846        Gaps: 24
  Percent Similarity: 48.734  Percent Identity: 27.215

alignment_block:
US-09-462-480-1 x T02345 ..
Align seg 1/1 to: T02345 from: 1 to: 1791

48 CGGCAACCCAGCCGCGCAGGAGCGCCGACATGGGCT.....GC 88
  :::::::::::::::::::::
465 ArgThrProSerArgArgSerArgSerGlySerProGlyLeuArgAs 481
  :::::::::::::::::::::
89 TCGGCACCATCCGCTGTGCAACCATCCGCTGGCTGGTGGATCA..... 132
  :::::::::::::::::::::
481 p.GlySerGlyThrProSerArgHisSerLeuSerGlySerSerProGly 497
  :::::::::::::::::::::
133 .....GGCCCCAGCGCGCGCGCGCTGCTGCTCGCGCGC 164
  :::::::::::::::::::::
498 MetLysAspIleProArgThrProSerArgGly.....ArgSe 510
  :::::::::::::::::::::
165 GGAG.....TCGCTACCTGGCGCAGGTGGGT 190
  :::::::::::::::::::::
510 rGluCysAspSerSerProGluProLysAlaLeuProGlnThrProArgp 527
  :::::::::::::::::::::
191 CGTTGACCCGCGCGCTGATGCTCAGCTGATCGAAAAAGCCGGTTGCC 240
  :::::::::::::::::::::
527 roArgSerArgSerProSerProGluLeuAsnAsnLysCysLeuThr 543
  :::::::::::::::::::::
241 CCCTCGGTGATGCGCGCGCTGTTGCGCGATCGTGGTGGTGGTGGCGC 290
  :::::::::::::::::::::
544 ProGln.....ArgGluArgSerGlySerGluSerValAspG1 557
  :::::::::::::::::::::
291 CGTCCCGGTGGTCCGGGAGCGATGGGCCAGGTTTCGCAATCCGCGGCT 340
  :::::::::::::::::::::
557 nLysThrValAlaArgThrProLeuGlyGlnArgSerArgSerGlySers 574
  :::::::::::::::::::::
341 CCACGACCGCGGTCTGCTGCGCGCGCACCGCTCGCGCAGGAGCGTAA 390
  :::::::::::::::::::::
574 erGlnGluLeuAsp...ValLysProSerAlaSerProGlnGlnArgSer 589
  :::::::::::::::::::::
391 GAAGACGACGAGGAGCGACTGGGACGAGAGGACGACTGGTGGCTCCGCT 440
  :::::::::::::::::::::
590 GluSer.....AspSerSerProAspSerLysAlaLysTh 601
  :::::::::::::::::::::
441 ANTGACAACAGACTTCCCGCGCACCCGCGCGGAAGACTTGCACACATT. 489
  :::::::::::::::::::::
601 rArgThrProLeuArgGlnArgSerArgSerGlySerProGluValA 618
  :::::::::::::::::::::
490 .....TTGGCGAGGAGGTAAGAGAGAAAGTAGTCCACGA 525
  :::::::::::::::::::::
618 spSerLysSerArgLysSerProArgArgSerArgSerGlySerPro 634
  :::::::::::::::::::::
526 TGGCAGAGATGAAGACCGCATCCGCTACCTCGGCGCAGGAGCAGGTAAT 575
  :::::::::::::::::::::
635 GluValLysAspLysProArgAlaAlaProArgAlaGlnSerGlySerAs 651
  :::::::::::::::::::::
576 TTCAGAGCGGATCTCGGCGCGCTGAAACCCAGATCG...ACCAGTGGA 622
  :::::::::::::::::::::
651 pSerSerProGluProLysAlaProAlaProArgAlaLeuProArgS 668
  :::::::::::::::::::::
623 GTCGACCGCAGGTTCGTTCCAGGCGCGCTGGCGCGCGCGGGGAGCG 672
  :::::::::::::::::::::
668 erArg.....SerGlySerSerLysGlyArg 677
  :::::::::::::::::::::
673 CCGCCGCGCGCGGTGGTGGCTTCCAAAGAG.....CAGCCAAT 713
  :::::::::::::::::::::
678 GlyProSerProGluGlySerSerSerThrGluSerSerProGluHisPr 694
  :::::::::::::::::::::
714 AAGCAGAGAGGAACTCGACGAGATCTCGACGAATATTCGTACGCGCG 763
  :::::::::::::::::::::
694 oProLysSerArgThrAlaArgArgGlySerArgSerProGluProL 711
  :::::::::::::::::::::
764 GGTCCAATACTCGA...GGCGCGCAGGAGCAGCAGCGCTGCTCT 810
  :::::::::::::::::::::
711 ysThrLysSerArgThrProProArgArgArgSerSerSerPro 727
  :::::::::::::::::::::
811 .....CGCAATGGGCTTC 824
```

```

Align_seg 1/1 to: Tl0032 from: 1 to: 100

525 ATGGCAGAGATGAAGACCGATGCGCTACCTCGGCAGGAGGCAGGTAA 574
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1 MetAlaGluMetIleThrGluAlaIleLeuThrGlnAlaAlaG1 17
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
575 TTTCGAGCGGATCTCGCGGCACCTGAAACCCAGATCGACGACGAGTGGAGT 624
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
17 nPheaspGlnIleAlaSerGlyLeuSerGlnGluArgAsnPheValAspS 34
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
625 GCACGGCAGGTTCTGTTGCAGGGCCAGTGGCGCGCGCGGGGAGCGGCC 674
||  ::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
34 erIleGlycInSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAla 50
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
675 GCCACGGCGCGGTGGTGCCTCCAAAGAGGACGACCAATAGCAGGAACA 724
||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
51 AlaLeuGlyAlaLeuGlyArgPheaspGluAlaMetGlnaspGlnIleAr 67
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
725 GGAACTCGACGAGATCTCGACGAATATTCTGTCAGCGCGCGCTCCAACTACT 774
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
67 gclnLeuGluSerIleValAspIysLeuAsnAsfSerGlyGlyAsnfyrt 84
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
775 CGAGGCGGCACGAGGAGCAGCAGCGCGCTGCTCGCAAAATGGCGTTC 824
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
84 hrLysThrAspAspGluAlaAsnGlnLeuLeuSerLysMetAsnPhe 100

seq_name: pir2:Tl0031
seq_documentation_block:
early secretory antigen target 6 protein homolog - Mycobacterium
C:Species: Mycobacterium leprae
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change
C:Accession: Tl0031
R:Eiglmelzer, K.; Honore, N.; Woods, S.A.; Caudron, B.; Cole, S.T.
Mol. Microbiol. 7, 197-206, 1993
A>Title: Use of an ordered cosmid library to deduce the genomic
A:Reference number: 216917: MUID: 93188700

```

```
alignment_scores:
  Quality: 186.00      Length: 91
  Ratio: 2.447         Gaps: 0
  Percent Similarity: 83.516      Percent Identity: 36.264

alignment_block:
  US-09-462-480-1 x T10031      ..
```

[illegible]

98 proGluprometSerAlaLeuAlaSerArqSerAlaProAlaMetGlnSe 114

```
alignment_block:
```

```

alignment_scores:
  Quality: 182.50      Length: 455
  Ratio: 1.031        Gaps: 28
  Percent Similarity: 38.901      Percent Identity: 27.473

alignment_block:
  US-09-463-480-1/rev x S53796
  ..

```

Align seg 1/1 to: S52796 from: 1 to: 403

```
1268 GCGGCATGACAACTCTCAGAGTCGCTCAAAACGTATATAACACGAGAAA 1219
|||||
15 AlaGlyMetProLysLeuArgSerThrAlaAsnArgAspAsnAspSerG1 31
|||||
1218 GGGCGAG.....ACGACGGAAGGTGCAACTGCGCCG 1187
|||||
31 yGlySerArgProProLeuLeuProProGlyGlyArgSerThrSerAlaL 48
|||||
1186 ATCCGCTGTTTCGCTATTCTAGCGAACTCGCGTTCCTATGCGGAACA 1137
|||||
48 ysPro...PheSerProProSerGlyProGlyArgPheProValProSer 63
|||||
1136 TCCAGTGCAGTTCCTTCGCGTTCGAGGACATTGCTGACCGCTTCGCTG 1087
|||||
64 ProGlyHisArg.....SerGlyProProGluProGluArgAsnArg... 77
|||||
1086 ATCGTCGCGCGCAGGTTCTGACGCGGTGTTTCAGCTCGGTAGCGGTGC 1037
|||||
78 .....MetPro.....ProProA 82
|||||
1036 GTCCCATTTTCCTGGACACCTGCTGATG.....CCTCCGAACGCTAC 993
|||||
82 rgProAspValGlySerLysProAspSerIleProProValProSer 98
|||||
992 CGCCCGAGCGCTCGCA...GCTTGGTCAGGACTGCTTCCCTTCGTCA 946
|||||
99 ThrProArgProIleGlnSerSerLeuHisAsnArgGlySerProProva 115
|||||
945 AGGAGGAATGAATGACGTGACATTTCCCTGGATTGCGTTCGCGCGG. 897
|||||
115 lProgly.....GlyProArgGlnProSerP 124
|||||
896 .....CCTCATACCGCGCAATTCACACTGCTGCTGTCA 861
|||||
124 roGlyProThrProProProPheProGlyAsnArgGlyThrAlaLeuGly 140
|||||
860 TGTTTTTCCTCGTTCCTTTCGTTATAGCGGTCAGAAAGCCATTTGCG 811
|||||
141 GlyGlySerIleArgGlnSerProLeuSerSerSer...SerProPheSer 156
|||||
810 AGGACAGCGCTGCT.....GCTGC 791
|||||
157 AsnArgProProLeuProProThrProSerArgAlaLeuAspAspLysPr 173
|||||
790 TCCTCGTCGCGCTCGAGTATTGGACGCGCGCTGACGAATATTCGTGCA 741
|||||
173 oProProProProProProValGlyAsnArgPro.....SerIleH 187
|||||
740 GATCTCGTCGAGTTCCTCTCTGCTTATTGGCTGCTTCTTGAAGCGCA 691
|||||
187 isArgGluAlaValPro..... 192
|||||
690 CCACGCGCGCTGGCGCGCTCCCGCGCGCGCGCTGCGCTGCGCTGC 641
|||||
193 ProProProGlnAsnAsnLysProPro..... 202
|||||
640 AACGAACCTGCGCTGAGTTCACCTGCTGATCTGGGTTTTCAGGTGCGC 591
|||||
203 .....ValProSerThrPro.....A 208
|||||
590 GGAGATCGCT.....CGAAATACCTGCTCCTGCGCGCGCGGTAGCGG 547
|||||
208 rgProSerAlaProHisArgProHisLeuArgPro...Pro..... 220
|||||
546 CATCGGTCTTCATCTCTGCCATGTGCTGACTACTTCTCTCTTTACCTTC 497
|||||
221 .....ProPro 222
|||||
496 TCGCAAAATGTTGGCAAGTCTTCGCGCGCGGTGGCGGAAGTCTGTT 447
|||||
```

```
223 Ser.....ArgProGlyProPro..... 228
446 GTCAATTACGGAGCTCACAGTCGCTCTCTTCCTCCAGTCGT..... 404
|||||
229 .....ProLeuProProSerSerSerGlyA 237
403 .....CCTCGCTGCTTCTTCACGCTCCTCGCGCA 374
|||||
237 snAspGluThrProArgLeuProGlnArgAsnLeu...SerLeuSerSer 252
|||||
373 CGGTGCGCGCGCGCACAGACCGCGCTGGTGGAGCCCGCGGATTGCGAA 324
|||||
253 SerThrProProLeuProSerProGly.....Ar 262
|||||
323 CCTTGGCCATCGCTCCCGGACCCACCGAGCGCGCGCACCGCTCACCGA 274
|||||
262 gSerGlyPro...LeuProProValProSerGluArgProProProp 278
|||||
273 CGATCCGCGCACAGCGCGCGCATCA.....CCGAGGGGGCAACCG 233
|||||
278 roValArgAspProProGlyArgSerGlyProLeuProProProPro 294
|||||
232 GCTTTTCATCAGCTGACATCAGCGCGCTGCGGTCAACGACCCACCT 183
|||||
295 ValSerArgAsnGlySerThrSerArgAlaLeuProAlaThr...ProGln 310
|||||
182 GCGCA.....GGTAGCGACTCCCGCGCGCAGCGCGCGCGCGCGC 142
|||||
311 LeuProSerArgSerGlyValAspSerProArgSerGlyProArgProPr 327
|||||
141 GCTGGGCGCTGATCCACCGACGCGGATGTTTCGACAGCGGACTGGTGC 92
|||||
327 oLeuProProAspArgPro..... 333
|||||
91 CGACGAGCCCATCTGCGCGCTTCTCTGCTGCGGTGGTTCGCGCGCGG 42
|||||
334 .....SerAlaGlyAlaProProPro 340
|||||
41 GTGCGCGCCACC 30
|||||
341 ProProProSer 344
|||||
seq_name: pir2:T02345
seq_documentation_block:
hypothetical protein KIAA0324 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C:Accession: T02345
R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.
re. J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A:Description: Sequencing of human chromosome 16p13.3.
A:Reference number: Z14664
A:Accession: T02345
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1791 <RIC>
A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650
C:Genetics:
A:Map position: 16
A:Introns: 1610/2; 1706/2
A>Note: KIAA0324
```

alignment_scores:
Quality: 179.50 Length: 493
Ratio: 0.820 Gaps: 25
Percent Similarity: 44.422 Percent Identity: 26.572

alignment_block:
US-09-462-480-1/rev x T02345

Align seg 1/1 to: T02345 from: 1 to: 1791

```
1271 TCGGCGCGCATGACAACTCTCAGAGTCGGCTCAAAAGCTATAAACACGAG 1222
      ::::::::::::::::::::: ||| |||
1269 AlalaAlaMetAsnLeuAlaSerAlaArgThrProAlaIleProThrAl 1285
      ::::::::::::::::::::: ||| |||
1221 AAAGGCGAGACCGGAGCGGAAGTGAAGTCCGCCGATCCCGTGTTCGCT 1172
      ::::::::::::::::::::: ||| |||
1285 aValAsnLeuAlaAspSerArgThrProAlaAlaAlaAlaMetAsnL 1302
      ::::::::::::::::::::: ||| |||
1171 AT...TCTACGGGAAGTCTGGGCTGCCCTATGCGAACATCCAGTCAGCT 1125
      ||| ||||| ::::::::::: ||
1302 euAlaSerProArgThrAlaValAla.....ProSerAlaVa 1314
      ::::::::::::::::::::: ||| |||
1124 TGCCTTCGGTGAAGCCATTGCTACCGGCTTCGCTGATCGTCGCCGCC 1075
      ::::::::::::::::::::: ||| |||
1314 l.....AsnLeuAlaAsp...ProArgT 1321
      ::::::::::::::::::::: ||| |||
1074 AGTTTCTCAGCGCGGTGTTTTCAGTCTGGTAGCGGTGCCATTTTGT 1025
      ::::::::::::::::::::: ||| |||
1321 hrProThrAlaProAlaValAsnLeuAlaGlyAlaArgThrProAlaAla 1337
      ::::::::::::::::::::: ||| |||
1024 CTGGACACCTGGTACGCTCGGAACCGCTACCGCCCGCCGCGCTGCGA 975
      ||| :::::: ||| :::::: |||
1338 LeuAlaAlaLeuSerLeuThrGlySerGlyThr.ProProThrAlaAlaA 1354
      ::::::::::::::::::::: ||| |||
974 GCTTGGTCAGGACTGCTTCCCTCGTCAAGGAGGGAATGAATGACGCTG 925
      ::::::::::::::::::::: ||| |||
1354 sn.....TyrProSerSerSerArgThrProGln..... 1363
      ::::::::::::::::::::: ||| |||
924 ACATTTCCCTGGATTGCGCTGTCGCCGCGCTCGATA.....CCGCGGAA 881
      ||| :::::: ||| :::::: |||
1364 .....AlaProAlaSerAlaAsnLeuValGlyProArgSe 1375
      ::::::::::::::::::::: ||| |||
880 ATTCCACTGCTCTCTGATCACTTTTTCGCTCGGTTCTTTTCGTATTACG 831
      ||| :::::: ||| :::::: |||
1375 rAlaHisAlaThrAla.....P 1381
      ::::::::::::::::::::: ||| |||
830 GGGTCAGAAGCCATTTCGAGGACGAGCGCTGCTGCTCTCTCGTCGG 781
      ||| :::::: ||| :::::: |||
1381 roValAsnIleAlaGlySerArgThrAlaAlaAlaLeuAlaProAlaSer 1397
      ::::::::::::::::::::: ||| |||
780 CCTCGAGTATGACGGCGGCT.....GACGAATATTCGTCGAGATC 737
      ||| :::::: ||| :::::: |||
1398 LeuThrSerAlaArgMetAlaProAlaLeuSerGlyAlaAsnLeuThrSe 1414
      ::::::::::::::::::::: ||| |||
736 TCGTCGAGTCTCTGCTTCTGCTTATGCTGCTTCTTGA....ACGCGAC 690
      ||| ||||| ||||| ||||| |||
1414 rProArgValProLeuSerAlaIleArgValSerGlyArgThrSerP 1431
      ::::::::::::::::::::: ||| |||
689 CACCG.....CGGCTGGCGCGCGCTCC 667
      ||| ||||| ||||| |||||
1431 roProLeuLeuAspArgAlaArgSerArgThrProProSerAlaProSer 1447
      ::::::::::::::::::::: ||| |||
666 CCGCGCGCG.....CGCGCCACTGCCCTGCAACGAACTGCC 629
      ||| ||||| ::::::::::: |||
1448 GlnSerArgMetThrSerGluArgAlaProSerProSerSerArgMetG 1464
      ::::::::::::::::::::: ||| |||
628 GTCGACTCCA..... 619
      ::::::::::::::::::::: ||| |||
1464 yGlnAlaProSerGlnSerLeuLeuProProAlaGlnAspGlnProArgS 1481
      ::::::::::::::::::::: ||| |||
618 ..CCTGGTCTGATCTGGTTTTCAGTTCGGCGGAGATCCGCTCGAAATTAC 571
      ||| ||||| ||||| ||||| |||
1481 erProValProSerAlaPheSerAspGlnSerArgCysLeuIleAlaGln 1497
      ::::::::::::::::::::: ||| |||
570 CTGCTCTCTCCCGAGGAGTACGGATCGGCTCTTCATCTCTGCCATGCTG 521
      ||| :::::: ||| :::::: |||
1498 ThrThrProValAlaGlySerGlnSerLeuSerSerGlyAlaValAlaTh 1514
      ::::::::::::::::::::: ||| |||
520 GACTACTTCTCTCTT..... 505
      ||||| |||
1514 rThrThrSerSerAlaGlyAspHisAsnGlyMetLeuSerValProAlaP 1531
      ||||| |||
```

```
504 ..TACCTTCTCGCCAAATGTTGGAAAGTCTTCGGCCCGGGTGGCCGG 457
      ::::::::::::::::::::: |||||
1531 roGlyValProHisSerAspValGlyGluPro.ProAlaSer..... 1544
      ::::::::::::::::::::: ||| |||
456 GAAGTCTCTTCTATTACGGGAGCTCACAG..... 426
      ||||| ||||| :::::: |||
1545 .....ThrGlyAlaGlnProSerAlaLeuAlaLaL 1556
      ::::::::::::::::::::: ||| |||
425 .....TCGTCC 420
      ::::::::::::::::::::: |||||
1556 euGlnProAlaLysGluArgArgSerSerSerSerSerSerSerSer 1572
      ::::::::::::::::::::: |||||
419 TCTTCGTCCTCCAGTCTCTCGCTCTCTTCTTTCACGCTCTCGCGCAGCG 370
      ||||| ||||| ::::::::::: |||
1573 SerSerSerSerSerSerSerSerSerSerSerSerSerSerGlu 1589
      ::::::::::::::::::::: ||| |||
369 TCCGCGCGCGACACCGCGGCTGGTGAGCCGCCGATTCGGAACCTT 320
      ::::::::::::::::::::: ||| |||
1589 ySerSerSerSer.Asp.....SerGlu 1596
      ::::::::::::::::::::: ||| |||
319 GCGCCATCGCTCCCGGACCCACCGGAG.....CGGCGCCACCGCTC 279
      ||| ||||| ||||| |||||
1597 GlySerSerLeuProValGlnProGluValAlaLeuLysArgValProSe 1613
      ::::::::::::::::::::: ||| |||
278 ACCGAGCATCCGGCAACAGCCCGCGCATCACGAGGGGGCAACCGGCTT 229
      ||||| ||||| :::::: |||
1613 rProThr..ProAlaProLysGluAlaValArgGluGlyArg..... 1626
      ::::::::::::::::::::: ||| |||
228 TTCGATCAGCTGAGACATCAGCGGCGTGGGTCAACGACCCACCTCGCG 179
      ||||| ||||| |||||
1627 .....ProProGluP 1630
      ::::::::::::::::::::: ||| |||
178 CA.....GGTAGCGACTCCGCGCGCAGCAGG 153
      |||
1630 roThrProAlaLysArgLysArgArgSerSerSerSerSerSerSer 1646
      ::::::::::::::::::::: ||| |||
152 CCGCGCGCGCTGGGCTGATCACCGACGCGGATGTTGACAG 103
      ::::::::::::::::::::: ||| |||
1647 SerSerSerSerSerSerSerSerSerSerSerSerSerSe 1663
      ::::::::::::::::::::: ||| |||
102 CGGACTGGTGGCGAGCAGGCCCATCTGCGCGGCTTCTCTGCTGGCTGGT 53
      ::::::::::::::::::::: ||| |||
1663 rSerSerSerSerSerSerSerSerSerSerSerSerSerS 1680
      ::::::::::::::::::::: ||| |||
52 TCGCGCGCGCGTGGCGGCC 33
      ||| ||||| |||
1680 erProSerProAlaLysPro 1686
      ::::::::::::::::::::: ||| |||
```

seq_name: pir2:B40505

seq_documentation_block:

hypothetical protein - suid herpesvirus 1 (strain Indiana-Funkhuser or Becker)

C/Species: suid herpesvirus 1

C/Date: 10-Apr-1992 #sequence_revision 10-Apr-1992 #text_change 02-Sep-2000

C/Accession: B40505

R/Cheung, A.K.

J. Virol. 65, 5260-5271, 1991

A/Title: Cloning of the latency gene and the early protein 0 gene of pseudorabies vi

A/Reference number: A40505; MUID:91374576

A/Accession: B40505

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1958 <CHE>

A/Cross-references: CB:M57505; NID:g334066; PIDN:AAA47468.1; PID:g334068

C/Superfamily: pseudorabies virus 1 nuclear antigen

alignment_scores:

Quality: 178.50 Length: 508

Ratio: 0.776 Gaps: 32

Percent Similarity: 45.276 Percent Identity: 28.346

alignment_block:

US-09-462-480-1 x B40505

Align seg 1/1 to: B40505 from: 1 to: 1958

```
31 GTGGCGGCGCACCGGC...GGCGGCAACCCA..... 57
|||||
1053 ValGlyGlyGluGlyArgGlyGlyGlyProArgValGlyLeuAlaG1 1069
58 .....GCCGACGAGGAAGCGCGCAGATGGCTGTCTGCGCACCATGTC 100
|||||
1069 yArgAspAlaAlaGluAlaAlaValAlcIyArgGlyValLeuGlyHisGlyP 1086
|||||
101 CG...CTGTGCAACCATCG.....CTGGCTGGTGA..... 129
||
1086 roGluArgAlaProGluProValValLeuGlyGlyGlyGlyGlyGly 1102
130 .....TCAGCCCGC...AGCGCGGG 146
1103 GlyGlyHisGluArgGlySerGlyValArgSerGlyProGluSerGluG1 1119
147 CGCGGCGCTG..... 156
1119 yAlaAlaLeuAlaProGlyProProValLeuPheValValAlaValAlav 1136
157 .....CTGCGCGCGAG.....TCG 171
||
1136 aAlaValProAlaGluGlyArgAlaGlyGluProLeuValLeuAla 1152
172 CTACCTGCGCGAGGTGGTGTGACCGCGCGCTGATGTCACGT 221
|||||
1153 ValProGlyAlaAlaGly.....ProGlyArgAlaAlaLe 1164
222 GATCGAAAGCGGTGCGCCCTCGTGATGCGCGCGCTGTTCGCGAT 271
||
1164 uLeuLeuAlaProLeuGlyArgTrpValArg...AlaGlyGlyGlyGly 1180
272 CGTGGTGACGGGTGGCGCGCTCGGTGGTGGTGGCGGAGCG..... 312
|||||
1180 laGlyValAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1196
313 .....ATGGCGCAGGGTTCGCAATCCGGCGGCTCCAC.CAGCCGCGTCC 355
||
1197 LeuGlyAlaGlyAlaGlyLeuGlyAlaGlyAlaGlyGlyProGlyAl 1213
356 TGGTCGCGCGCGCACCGCTCGCGCAGGAGCGTGAAGAAGACGA..... 398
|||||
1213 aGlyGluAlaGly.....GlyGlyAlaArgArgArgArgArg 1227
399 .....CGAGGACGACTGGGA 413
1227 rgTrpAspGluAlaGlyLeuLeuGlyProGluArgGlyGluAlaGly 1243
414 CGA.....AGAGCAGACTGGTGAGCTCCCGTAA 442
||
1244 ArgGlyLeuArgGlyProGlyProArgGlyGlyLeuGlyGluProGly.. 1259
443 TGACAACAGACTTCGCGCGCACCGCGCGGCGGAAGACTTGCCCAACATTTTG 492
|||||
1260 .....ProGlyHisValGlyArg.....G 1266
493 CGGAGGAAGGTAA..... 506
1266 lyGluGluGlyArgGlyValGlyProGlyGlyLeuAlaGlyAlaGlyPro 1282
507 .....GAGAGAAGTAGTCCACCATGGCAGAGATGAA 538
|||||
1283 ValHisAlaAlaHisGluArgHisGlyAla..GlyAspGluGly 1298
539 GACCGATGCCCTACCTCGCGGCGAGGAGGAGTAATTTTCGAGCGGATCT 588
|||||
1299 AspArgValArgGlyLeuLeuProLeuGlyArgAlaGlyProGlyAsp.. 1314
```

```
589 CGGCGGACCTGAAACCCAGATCGACACCGAGGTGGAGTCTGACGCGCAGGTTCG 638
|||
1315 ArgValAlaGluArgGluGlnArgGlyGlyHisLeu.....LeuGluA 1329
639 TTGACAGGCCAGTGGCGCGCGC.....GGCGGAGCGCGCGCCAGCG 682
|||||
1329 laGlyGlyProGluGlyArgGlyAlaGlyGlyArgGlyGlnProGlu 1345
683 CGCGGTGGT.....CGCTTCCAAGAGCAGCAATAGCAGAGCAGG 726
|||||
1346 ArgAlaGlyGlnGlnAlaLeuGluAspAlaAlaGlyGlnAspAlaG1 1362
727 AACTCGACGAGATCTCGAGAATATTCGTAGCGCGGCTCCAATCTCTG 776
|||||
1362 yValArgGln...LeuAlaGlyHisAlaAlaGlyLeuArg.....G 1375
777 AGGCGCGGAGGAGCAGCAGCAGCGCTGTCTCGCAATGGCTTCG 826
|||||
1375 lyGlyGluGlyGlyAlaAspAlaGlyAlaGluGlyLeuAspGlyArgLeu 1391
827 ACCCGCTAATACGAAAGAAAGCGGAGCAAAACATGACAGACGACGATG 876
1392 ProGly..... 1393
877 GAATTCGCGGTATCGAGCGCGCGCAATCCAGGAAATGTCA 926
|||||
1394 .....AlaGlyValArgGlyAlaAlaArgValGlyHisValGlyValG 1408
927 CGTCCATTTCCTCTCTGACGAGGAGGAGCAGTCCCTGACCAAGCTC 976
|||||
1408 lyProAlaGluAlaLeuGlnAspGluGly.....LeuLeuGlyAlaAla 1422
977 GCAGCGCGCTGGCGGTAGCGGTTCGAGGCGGTACC..... 1013
1423 ValAlaAlaAlaHisGlyHisGlyAla..HisArgValArgGlnGlyProG 1439
1014 .....AGGGTCTCCAGCAAAATGGCGCGCGCTACCGAGCTGAACAACG 1061
|||||
1439 luArgValLeuGlyGlyHisGlyValProAspVal..... 1450
1062 CGCTGCAGAACCTGGCGCGGAGGATCAGCGAGCGCGGTACGCAATGGCT 1111
|||
1451 .....ArgGlnArgArgGlyHisAlaAlaGlyAsp..... 1460
1112 TCGACCGAAGCAACGCTCACTGGGATGTCATAGGCGCAACGCGCGAT 1161
|||||
1461 .....GluGlyAlaValAlaValGly.....ArgValAspProAlaL 1473
1162 CGCCTAGATAGCGAAACACGGGATCGCGGAGTTCGACCTTCGCTC... 1208
|||||
1473 euAlaGluLeuValGluAlaLeuValGlyGlyLeuHisProArgValLeu 1489
1209 .....GGTCTCGCC 1217
1490 GlnProGlyHisGlyLeuAla 1496
seq_name: pir2:T08179
```

seq_documentation_block:

LRG5 protein - Chlamydomonas reinhardtii

C:Species: Chlamydomonas reinhardtii

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999

C:Accession: T08179

R:Gloeckner, G.; Beck, C.F.

submitted to the EMBL Data Library, October 1996

A:Description: Molecular characterization of a gene (LRG5) involved in blue light sig

A:Reference number: Z16399

A:Accession: T08179

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-640 <GLO>

A:Cross-references: EMBL:U73817; NID:g1644369; PID:g1644370

C:Genetics:

A:Title: cDNA cloning and chromosomal mapping of the mouse type VII collagen gene (Col7A)

A:Reference number: A45748; MUID:93315168

A:Accession: A45748

A>Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-920 <L11>

A:Cross-references: GB:S63654; NID:g386656; PIDN:AAB27492.1; PID:g386657

A:Experimental source: epidermal keratinocyte

A>Note: sequence extracted from NCBI backbone (NCBIN:135000, NCBIP:135001)

C:Superfamily: unassigned collagens

alignment_scores:

Quality:	174.00	Length:	421
Ratio:	0.911	Gaps:	24
Percent Similarity:	45.368	Percent Identity:	26.603

alignment_block:

US-09-462-480-1 x A45748 ..

Align seg 1/1 to: A45748 from: 1 to: 920

```
27 CCAGTGGCGGCGCCGCGCGCCACCCAGCCGACGAGGAGCGCGC 76
||||| : : : : : : : : : : : : : : : : : : : : : :
292 ProGlyProGlnGlyProAlaGlyArgProGlyGluLysGlu 308
77 AGAT.....: : : : : : : : : : : : : : : : : :
308 YAspCysGluAspGlyGlyProGlyLeuProGlyGlnProGlyPro. 324
91 GGCACAGTCGCGTGTGCAACCATCCGCTGGCTGGTATCAGCCCGCAG 140
||| : : : : : : : : : : : : : : : : : : : : : :
325 GlyGluProGlyLeuArgGlyAlaPro.....GlyMetThrGlyProLy 339
141 CGCGGCGCGCGCGCTGCTCCGCGCGGAGTCCTACCTGCGCAGGTGGT 190
: : : : : : : : : : : : : : : : : : : : : :
339 sGlyAspArgGlyLeuThrGlyThrProGlyGluProGlyValLysGly 356
191 CGTTGACCGCGCGCGCTGTGTCTCAGCTGATCGAAAGCCGTTGCC 240
: : : : : : : : : : : : : : : : : : : : : :
356 LuArgGlyHis.....ProGlyProValGly 364
241 CCCTCGGTGATCGCGCGCGGTGTCGCGGATCGTCGGTACGGTGGCGC 290
||| : : : : : : : : : : : : : : : : : : : : : :
365 ProGlnGlyLeuProGlyAlaAlaGlyHisProGlyValGluGlyProGl 381
291 CGCTCCGCGTGGTCCG.....: : : : : : : : : : : :
381 uGlyProProGlyProThrGlyArgGlyGluLysGlyGluProGlyA 398
307 .....: : : : : : : : : : : : : : : : : :
398 rgProGlyAspProAlaValGlyProGlyGlyAlaGlyAlaLysGlyGlu 414
322 GGTTCGCAATCCGC.....GGCTCCACC.. 345
: : : : : : : : : : : : : : : : : : : : : :
415 LysGlyGluAlaGlyLeuProGlyProArgGlyAlaSerGlySerLysG 431
346 .....AGCCCGGTCTGGTCGCGCGCACCCCTCGCGCAGGAGC 385
: : : : : : : : : : : : : : : : : : : : : :
431 yGluGlnGlyAlaProGlyLeuAlaLeuProGlyAspProGlyProLysG 448
386 GTGAAGAAGACGACGAGG.....ACGACTGGGACGAAGAGGAC 423
: : : : : : : : : : : : : : : : : : : : : :
448 LyAspProGlyAsp.ArgGlyProIleGlyLeuThrGlyArgAlaGlyPr 464
424 GACTGGTGAAGTCCCGTAATGACACAGACTTCCCGCCACCGCGCGG 473
||| : : : : : : : : : : : : : : : : : : : : : :
464 oThrGlyAsp.....SerGlyProProGlyGluL 474
474 AAGACTTGCCCAACATTTTGGCGAGGAAGTTAAAGAGAG.....AAGTA 517
: : : : : : : : : : : : : : : : : : : : : :
474 ySgGlyGluPro.....GlyArgProGlySerProGlyProVal 486
```

```
518 GTCCAGCATGCGCAGAGATGAAGACCGATCCGCGTACCCCTCGGGCAGGAG 567
: : : : : : : : : : : : : : : : : : : : : :
487 GlyProArgGlyArgAspGlyGlu.....AlaGlyGlu 497
568 CAGGTAAATTCGACGGATCTCCGGCGACCTGAAACCCAGATCCGACGAG 617
: : : : : : : : : : : : : : : : : : : : : :
497 u.....LysGlyAspGluGlyIleProGlyGluProGlyLeuProG 511
618 GTGGAGTCGACGCGGAGGTTCGTT.....GCAGGGGCA 649
|| : : : : : : : : : : : : : : : : : : : : : :
511 LyLysAlaGlyGluArgGlyLeuArgGlyAlaProGlyProArgGlyPro 527
650 GTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCTCC 699
||| : : : : : : : : : : : : : : : : : : : : : :
528 ValGlyGluLysGlyAspGlnGlyAspProGlyGluAspGly..... 541
700 AAGAAGCAGCAATAAGCAGAGCAGGAACTCGACGAGATCTCGACGAAT 749
||| : : : : : : : : : : : : : : : : : : : : : :
542 .ArgAsnGlySerProGlySerSerGlyProLysGlyAspArgGlyGluP 558
750 ATTGCTCAGGCGCGCGCTCAATACTCGAGGCGCGCGAGGAGCAGCAGCA 799
: : : : : : : : : : : : : : : : : : : : : :
558 roGlyProProGlyProPro.....GlyArgLeuValAspAla 570
800 GGCCTGTCTCTCGCAAAATGGCTTCTGACCGCTAAATACGAAAGAAACG 849
||| : : : : : : : : : : : : : : : : : : : : : :
571 Gly.IleGluSerArgAspLysGlyGluProGlyGlnGluGlyProArgG 587
850 GAGCAAAACATGACAGACGACAGTGAATTCGCGGGGTATCGAGGC... 897
|| : : : : : : : : : : : : : : : : : : : : : :
587 lyProLysGlyAspProGlyProGlyValSer.GlyGluArgGlyI 603
898 .....CGCGCAAGCGCAATCCAGGCAATGTACCTCCATTTCAT 937
||| : : : : : : : : : : : : : : : : : : : : : :
603 eAspGlyLeuArgGlyProProGlyPro.....GlnGlyA 615
938 TCCCTCTCTGACGAGGGAGCA.....GTCCCTGACCAAGCTCGCAGC 981
||| : : : : : : : : : : : : : : : : : : : : : :
615 spProGlyValArgGlyProAlaGlyAspLysGlyAspArgGlyProPro 631
982 GGCCTG...GGCGGTAGCGGTTCGGAGGCTACCGAGGTGTCCAGCAAA 1028
||| : : : : : : : : : : : : : : : : : : : : : :
632 GlyLeuAspGlyArgSerGlyLeuAspGlyLysProGlyAlaProGlyPr 648
1029 AATGGGACCGCGCTACCGAGCTGAACAACGCGCTGCAGAACTGGCG 1078
||| : : : : : : : : : : : : : : : : : : : : : :
648 oProGlyLeuHisGly.....AlaSerGlyLysAlaGlyAspProGlyA 663
1079 CGGAC 1083
|||
663 rgAsp 664
```

OM of: US-09-462-480-1 to: SwissProt_40.* out_format : pfs
Date: Jul 22, 2002 1:41 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+nt2p.model -DEV=xlh
-O=/cgn2.1/USPOT_spool/US09462480/runat_18072002_164420_19600/app_query.fasta_1.2850
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
-CAPEXT=4.000 -MINMATCH=0.100 -LOPECL=0.000 -LOPEXT=0.000
-GAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -YGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480.ecgn1_1.29 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-462-480-1
Query length: 1277
Database: SwissProt_40.*
Database sequences: 105224
Database length: 38719550
Search time (sec): 49.760000

score_list:	Sequence	Strd Orig	zscore	EScore	Len	Documentation
	SwissProt_40:CF10_MYCTU	481.00	471.11	1.5e-18	99	! 069739 mycobacterium tubercul
	SwissProt_40:ES66_MYCTU	474.00	464.91	3.6e-18	94	! 057165 mycobacterium tubercul
	SwissProt_40:YHL1_EBV	212.50	204.07	0.0002	660	! P03181 Epstein-Barr virus (str
	SwissProt_40:WAP1_HUMAN	189.50	184.45	0.0028	503	! O43516 homo sapiens (human)
	SwissProt_40:ES66_MYCLE	186.00	193.64	0.0046	95	! O50206 mycobacterium leprae
	SwissProt_40:CF10_MYCLE	182.00	189.56	0.0074	99	! O33084 mycobacterium leprae
	SwissProt_40:VNUA_PRVKA	172.50	159.18	0.0208	1733	! P33485 pseudorabies virus (
	SwissProt_40:DRPL_HUMAN	170.50	160.15	0.0268	1183	! P54258 rattus norvegicus (rat
	SwissProt_40:IRS2_HUMAN	169.50	158.37	0.0301	1324	! Q9Y4H2 homo sapiens (human)
	SwissProt_40:T2D3_HUMAN	168.50	158.93	0.0343	1083	! O00268 homo sapiens (human)
	SwissProt_40:DRPL_HUMAN	165.00	154.96	0.0521	1185	! P54259 homo sapiens (human)
	SwissProt_40:CAFE_RIFPA	164.50	155.56	0.0557	1027	! P30754 riftia pachyptila (b
	SwissProt_40:CA21_BOVIN	164.50	153.44	0.0551	1364	! P02465 bos taurus (bovine)
	SwissProt_40:MY15_HUMAN	163.50	145.37	0.0598	3530	! Q9UKN7 homo sapiens (human)
	SwissProt_40:CA17_HUMAN	163.00	146.26	0.0640	2944	! Q02388 homo sapiens (human)
	SwissProt_40:EXLP_TOBAC	162.50	160.27	0.0734	426	! Q03211 nicotiana tabacum (c
	SwissProt_40:VRP1_YEAST	161.00	153.98	0.0857	817	! P37370 saccharomyces cerevis
	SwissProt_40:CA21_RANCA	160.50	149.72	0.0893	1355	! O42350 rana catesbeiana (bu
	SwissProt_40:IE18_PRVIF	160.50	149.16	0.0890	1461	! P11675 pseudorabies virus (
	SwissProt_40:CA21_MOUSE	159.50	148.69	0.1007	1372	! Q01149 mus musculus (mouse)
	SwissProt_40:D4DR_HUMAN	159.00	156.29	0.1116	467	! P21917 homo sapiens (human)
	SwissProt_40:IE18_PRVKA	158.50	147.35	0.1134	1446	! P33479 pseudorabies virus (
	SwissProt_40:CA25_HUMAN	158.50	147.10	0.1132	1496	! P05997 homo sapiens (human)
	SwissProt_40:GPI_CHURE	157.50	153.58	0.1328	555	! Q9FQ66 chlamydomonas reinhard
	SwissProt_40:ICP0_HSV2H	157.50	150.61	0.1308	825	! P28284 herpes simplex virus
	SwissProt_40:IE18_PRVKA	156.50	145.47	0.1443	1446	! P33479 pseudorabies virus (
	SwissProt_40:S3A2_MOUSE	156.00	153.33	0.1602	475	! P06203 mus musculus (mouse)
	SwissProt_40:CA21_RAT	156.00	145.39	0.1536	1372	! P02466 rattus norvegicus (rat
	SwissProt_40:CA11_CHICK	155.00	144.96	0.1533	1453	! P02457 gallus gallus (chick)
	SwissProt_40:S3A2_HUMAN	155.00	152.57	0.1809	464	! Q15428 homo sapiens (human)
	SwissProt_40:EXTN_TOBAC	155.00	150.40	0.1789	620	! P13983 nicotiana tabacum (c
	SwissProt_40:CA12_MOUSE	155.00	143.99	0.1729	1459	! P28481 mus musculus (mouse)
	SwissProt_40:SES_RAT	153.00	146.37	0.2252	825	! Q63003 rattus norvegicus (rat)
	SwissProt_40:CA21_HUMAN	153.00	142.60	0.2208	1366	! P08123 homo sapiens (human)
	SwissProt_40:IE18_PRVIF	153.00	142.09	0.2202	1461	! P11675 pseudorabies virus (
	SwissProt_40:M3K1_RAT	153.00	141.93	0.2200	1493	! Q62925 rattus norvegicus (rat
	SwissProt_40:GSRI_HUMAN	152.50	141.38	0.2336	1509	! Q9NZM4 homo sapiens (human)
	SwissProt_40:SPD1_NEPL	152.00	146.18	0.2551	747	! P19837 nephila clavipes (ort
	SwissProt_40:CA21_ONCMY	152.00	141.71	0.2492	1356	! Q93484 oncorhynchus mykiss
	SwissProt_40:ICP0_HSV11	151.00	144.96	0.2874	775	! P08393 herpes simplex virus

SwissProt_40:CALL_MOUSE	+	151.00	140.25	0.2804	1453	! P11087 mus musculus (mo
SwissProt_40:ICP0_HSV2H	+	150.50	144.02	0.3046	825	! P28284 herpes simplex vi
SwissProt_40:CA13_HUMAN	+	150.50	139.71	0.2977	1466	! P02461 homo sapiens (hu
SwissProt_40:CA1B_MOUSE	+	150.50	138.16	0.2953	1804	! Q61245 mus musculus (mo
SwissProt_40:CA13_BOVIN	+	150.00	141.75	0.3205	1049	! P04258 bos taurus (bovi
seq_name: SwissProt_40:CF10_MYCTU						
seq_documentation_block:						
ID	CF10_MYCTU	STANDARD;	PRT;	99	AA.	
AC	069739;					
DT	16-OCT-2001	(Rel. 40, Created)				
DT	16-OCT-2001	(Rel. 40, Last sequence update)				
DT	16-OCT-2001	(Rel. 40, Last annotation update)				
DE	10 kDa culture filtrate antigen cfp10.					
GN	CFP10 OR LHP OR RV3874 OR MT3988 OR MTV027.09.					
OS	Mycobacterium tuberculosis.					
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;					
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.					
OX	NCBI_TaxID=1773;					
RN	[1]					
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.					
RC	STRAIN=H37RV;					
RX	MEDLINE=99061212; PubMed=9846755;					
RA	Berthet F.X., Rasmussen P.B., Rosenkrands I., Andersen P.,					
RA	Gicquel B.;					
RT	"A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel					
RT	low-molecular-mass culture filtrate protein (CFP-10).";					
RL	Microbiology 144:3195-3203(1998).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=H37RV;					
RX	MEDLINE=98295987; PubMed=9634230;					
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,					
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekaita F.,					
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,					
RA	Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,					
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,					
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,					
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,					
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;					
RT	"Deciphering the biology of Mycobacterium tuberculosis from the					
RT	complete genome sequence.";					
RL	Nature 393:537-544(1998).					
[3]						
RP	SEQUENCE FROM N.A.					
RC	STRAIN=CDC 1551 / Oshkosh;					
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,					
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,					
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,					
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,					
RA	Bishai W.;					
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and					
RT	laboratory strains.";					
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.					
CC	-!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
CC	EMBL; AF004671; AAC83445.1; -					
DR	EMBL; AL022120; CAAL1966.1; -					
DR	EMBL; AE007190; AAK48356.1; -					
DR	TIGR; MT3988; -					
DR	TubercuList; RV3874; -					
KW	Complete proteome.					
FT	INIT_MET 0 0					
SQ	SEQUENCE 99 AA; 10663 MW; EBCAE6A996C5489D CRC64;					

alignment_scores:
Quality: 481.00 Length: 99
Ratio: 4.859 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.990
alignment_block:
US-09-462-480-1 x CF10_MYCTU ..
Align seg 1/1 to: CF10_MYCTU from: 1 to: 99

528 GCAGAGATGACGACGATCGCGCTACCTCGGCGGAGGAGGAGGAGTAATTT 577
|||||
1 AlagluMetlysthrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 17
|||||
578 CGAGCGGATCCGCGGACCTGAAACCCAGATCGACCGAGTGGAGTCGA 627
|||||
17 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerT 34
|||||
628 CGGCGAGTTCGTCAGGCGCAGTGGCGCGCGCGGCGGAGCGCGCGCC 677
|||||
34 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 50
|||||
678 CAGCGCGCGTGGTGGCTTCCAGAGAGCAGCCCAATAGCAGAGCAGGA 727
|||||
51 GlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 67
|||||
728 ACTCGAGAGATCGCAGGAAATATCTGACGCGCGGCTCAATACTCGA 777
|||||
67 uLeuAspGluSerThrAsnIleArgGlnAlaGlyValGlnTrpSerA 84
|||||
778 GGGCGGCGAGGAGCAGCAGCAGCGCTGCTCGCAATGGGCTTC 824
|||||
84 rgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 99
|||||

seq_name: SwissProt_40:ESA6_MYCTU

seq_documentation_block:
ID ESA6_MYCTU STANDARD; PRT; 94 AA.
AC Q57165; O84901;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 6 kDa early secretory antigenic target (ESAT-6).
GN ESAT6 OR RV3875 OR RV3989 OR RV027.10.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=95204931; PubMed=7897219;
RA Andersen P., Andersen A.B., Sorensen A.L., Nagai S.;
RT "Recall of long-lived immunity to Mycobacterium tuberculosis infection
in mice.";
RL J. Immunol. 154:3359-3372(1995).
RN [2]
RP SEQUENCE FROM N.A., SEQUENCE OF 1-10, AND CHARACTERIZATION.
RC SPECIES=M.tuberculosis; STRAIN=ERDMAN;
RX MEDLINE=95247251; PubMed=7729876;
RA Soerensen A.L., Nagai S., Houen G., Anderson P., Anderson A.B.;
RT "Purification and characterization of a low-molecular-mass T-cell
antigen secreted by Mycobacterium tuberculosis.";
RL Infect. Immun. 63:1710-1717(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeter K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis;
RX MEDLINE=96200095; PubMed=8631702;
RA Mahairas G.G., Sabo P.J., Hickey M.J., Singh D.C., Stover C.K.;
RT "Molecular analysis of genetic differences between Mycobacterium bovis
BCG and virulent M. bovis.";
RL J. Bacteriol. 178:1274-1282(1996).
RN [6]
RP SEQUENCE OF 1-70 FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=99061212; PubMed=9846755;
RA Berthet F.-X., Rasmussen P.B., Rosenkrands I., Andersen P.,
RA Gicquel B.;
RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel
low-molecular-mass culture filtrate protein (CFP-10).";
RL Microbiology 144:3195-3203(1998).
CC -!- FUNCTION: NOT KNOWN. ELICITS HIGH LEVEL OF INF-GAMMA FROM MEMORY
EFFECTOR CELLS DURING THE FIRST PHASE OF A PROTECTIVE IMMUNE
RESPONSE.
CC -!- SUBCELLULAR LOCATION: Secreted
CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U34848; AAC44033.1; -;
DR EMBL: X79562; CAAS56099.1; -;
DR EMBL: AL022120; CAAL17967.1; -;
DR EMBL: AF007190; AAK48357.1; -;
DR EMBL: AF004671; AAC83446.1; -;
DR TIGR: MT3989; -;
DR TuberculList; Rv3875; -;
KW Antigen; Complete proteome.
FT INIT_MET 0
SQ SEQUENCE 94 AA; 9773 MW; 19245B0EC478BC84 CRC64;

alignment_scores:
Quality: 474.00 Length: 94
Ratio: 5.043 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-480-1 x ESA6_MYCTU ..

Align seg 1/1 to: ESA6_MYCTU from: 1 to: 94


```

863 ACAGAGCAGCAGTGAATTTTCGGGTATCGAGCGCGCGCAAGCGCAAT 912
1 ThrGluGlnGlnTTPAsnPheAlaGlyLeuAlaAlaAlaSerAlaI 17
913 CCAGGGAAATGTCAGTCCATTATTCCTCTTCAGAGGGAGAGCAGT 962
17 eGlnGlyAsnValThrSerIleHisSerLeuLeuAspGluGlyLysGlns 34
963 CCCTGACCAAGCTCGAGCGCGGTGGCGGTAGCGTTCGGAGCGGTAC 1012
34 erLeuThrLysLeuAlaAlaAlaIlePheGlySerGlySerGluAlaIle 50
1013 CAGGTGTCCAGCAAAATGGACCCAGCTACCGAGCTGAACAAGC 1062
51 GlnGlyValGlnGlnLysTyrAspAlaThrAlaThrGluLeuAsnAla 67
1063 GCTCAGAACCTGCGCGGAGCAGTACGGAAGCGCGGTTCAGCAATGCTT 1112
67 aLeuGlnAsnLeuAlaAlaThrIleSerGluAlaGlyGlnAlaMetAla 84
1113 CGACGAGGCAAGCTCACTGGGATGTTGCA 1144
84 erThrGluGlyAsnValThrGlyMetPheAla 94

```

seq_name: SwissProt_40:YHL1_EBV

seq_documentation_block:

```

AC YHL1_EBV STANDARD; PRT; 660 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLF1 protein.
OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01555; ; NOT_ANNOTATED_CDS.
DR PIR; A03742; Q0BE3.
KW Hypothetical protein; Early protein; Repeat.
FT DOMAIN 149 648 4 x 125 AA TANDEM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

```

alignment_scores:

```

Quality: 212.50 Length: 560
Ratio: 0.944 Gaps: 29
Percent Similarity: 40.179 Percent Identity: 25.000

```

alignment_block:

US-09-462-480-1/rev x YHL1_EBV

```

Align seg 1/1 to: YHL1_EBV from: 1 to: 660
1202 AGGTGCAACTGCGCGATCCCGTGTTT .....CGCTATTCTACGCGAAC 1159
136 ArgGlyProArgProHisProAlaPheGlnValGlnTrpSerAlaArgAs 152
1158 TCGCGGTTGCCTATGCGAACATCCAGTGACGTTGCTTCCGTGCGTCAAGC 1109
152 nProGlyCysPro...ArgThrTrpArgArgArgSerGlyAlaGlnArg 168
1108 CATGTCCTGACCGGCTTCGCTGATCGTCCGCCAGGTTCTGCACCGGCT 1059
168 lyHisPro.....ProProGlyAlaGlyGlnArg 177
1058 TGTTGAGCTCGGTAGCGGTGGCTCCCATTTTGTCTGGACACACCTGGTAC 1009
178 ProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyTh 194
1008 GCCT.....CCGAACCGCTACC 992
194 rProAlaAlaProGlyProGlyGlyAlaAlaValProSerGlyAlaI 211
991 GCCCGAGCGCGTGGGACGTTGTCAGGACTGCTTCCCTCCCTCGTCAAGA 942
211 hrProHisProGluArgGly...SerGlyProAlaAspProProAlaAla 226
941 GGAATGAATGACGTGACATTTCCCTGGATTGCGCTTCCG..... 900
227 Ala.....ArgLeuProProGluAr 233
899 ....CGGCTCGATACCGCGAAATTCACCTGCTCTGTCATGTTTTT 854
233 gGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysProA 250
853 GCTCGGTTCTTTTCGTATTAGCGGTGAGAGCCCATTT..CCGAGGACA 805
250 laGlyProProThrArgSerGlyAlaAlaGlnArgThrHisArg 266
804 GCGCTGCTGCTGCTCTCTCGTCGG..... 781
267 ArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArgTh 283
780 .....CCCTCGAGTATTG 768
283 rTrpArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaG 300
767 GACCGCGCGCTGACGAATATTCGTGAGATCTCGTCGAGTTCCTGCTTCT 718
300 lyGlnArgPro.....SerGlyProThrGlyGlyArgProAlaAla 313
717 GCTTATTGGCTGCTTCTTGGAGCGCACACCGCGCTGGCGCGCGCTC 668
314 Pro.....GlyAlaProGlyThrProAlaAlaProGlu 324
667 CCG.....CGCGCC 657
324 yProGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluA 341
656 GCGCCACTGGCCCT.....GCAACGAACCTGCGC..... 628
341 rgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlu 357
628 ..... 628
358 ArgGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysPr 374
627 ....TCGACTCCACCTGCTGATCTGGTTCAGTTCAGTCCCGG..... 589
374 oAlaGlyProProProThrArgSerGlyAlaAlaGlnArgThrHisA 391
588 .....AGATCCGCTCGAAT..... 574
391 rgArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArg 407

```

```
573 .....TACCTGCCCTCCGCCG 557
    ::::|::|::|::|
408 ThrTrpArgArgSerGlyAlaGlnArgGlyHisProProGlyAla 424
    ::::|::|::|::|
556 AGGTAGCGGATCGTCTTCATCTCCGACCTGGACTACTTCTCTC 507
    ||| ||| ||| |||
424 aGlyGlnArg.....ProSerGlyProThrGlyGlyA 435
506 TTTACCTCTCCGCCAAAATGTGGCAAGTCTTCGGCGCCGGT...GCG 460
435 rgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGly 451
459 CGGAAGTCTGTGTCAATACGGAGCT.....CACGACGCTCTCTT 416
    ||| ::::|::|::|::| ||| |||
452 GlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlyS 468
415 GTGCCAGTCTCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 384
    |||::|::|::|::|::|::|::|::|
468 erGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGln 484
383 .....TCTCGCGCGAGCGG 370
485 GluProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaG 501
369 TCGCGGCGCGACACCGCGCTGGTG.....GAGC 338
    ||| ||| ||| ::|
501 yProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgP 518
337 CGCGGATGTCACCTCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 296
    ||| ||| ||| ||| ||| |||
518 roProGlyCysProArgSerAlaArgAsnProGly.CysProArgThr 534
295 .....GAGCGCGCGCACCGCTCACCGACGATCGCGC 265
534 pArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyG 551
264 AACAGCCGCGGATCACCGAGGGGCAACCGGCTTTTCGATCAGCTGAG 215
    |||::|::|::|::|::|::|::|::|
551 lnArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaPro 567
214 ACATCAGCGGCGTGC.....GGTCAACGACCCACCTGCGCGCAGGT 174
    ::|::|::|::|::|::|::|::|
568 GlyThrProAlaAlaProGlyProGlyGlyGlyAlaAlaValProSerG 584
173 AGCGACTCGG.....CGCGCAGCAGGCGCG..... 149
584 yAlaThrProHisProGluArgGlySerGlyProAlaAlaAspProAla 601
148 .....CGCCGCGCTGGGCGCTGATCCAC 125
601 laAlaArgLeuProProGluArgGlnGluProArg..... 612
124 CAGCCAGCGGATGTTCCACAGCGGCTGTCGCGCAGCGCCCA..... 80
    ||| ||| ::::|::|::|::| ||| |||
613 LeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProPro 629
79 .....TCTGCGCGGCTCTCTGCTGCGG...CTGGGTT 52
629 oThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyC 646
51 GCGCGCGCGGCTGCGCGCCACCTGCG 26
    ||| ||| ::::|::|::|::| ||| |||
646 ysProArgSerAlaArgAsnProGly 654
```

seq_name: SwissProt_40:WAIP_HUMAN

seq_documentation_block:

```
ID WAIP_HUMAN STANDARD; PRT; 503 AA.
AC Q43516; Q9UNP1; Q15220;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE Wiskott-Aldrich syndrome protein interacting protein (WASP interacting
GN protein) (PRL-2 protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98070810; PubMed=9405671;
RA Ramesh N., Anton I.M., Hartwig J.H., Geha R.S.;
RT "WIP, a protein associated with Wiskott-Aldrich syndrome protein,
RT induces actin polymerization and redistribution in lymphoid cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:14671-14676(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=tonsil;
RA Kreideweiss S., Delany-Heiken P., Nordheim A., Ruhlmann A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 192-503 FROM N.A., AND VARIANT ALA-495.
RX MEDLINE=99218549; PubMed=10202051;
RA Stewart D.M., Tian L., Nelson D.L.;
RT "Mutations that cause the Wiskott-Aldrich syndrome impair the
RT interaction of Wiskott-Aldrich syndrome protein (WASP) with WASP
RT interacting protein.";
RL J. Immunol. 162:5019-5024(1999).
CC -1- FUNCTION: MAY HAVE DIRECT ACTIVITY ON THE ACTIN CYTOSKELETON.
CC INDUCES ACTIN POLYMERIZATION AND REDISTRIBUTION.
CC -1- SUBUNIT: BINDS TO WASP, PROFILIN AND ACTIN.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PERIPHERAL BLOOD
CC MONONUCLEAR CELLS, SPLEEN, PLACENTA, SMALL INTESTIN, COLON,
CC THYMUS. LOWER EXPRESSION IN OVARY, HEART, BRAIN, LUNG, LIVER,
CC SKELETAL MUSCLE, KIDNEY, PANCREAS, PROSTATE, TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF031588; AAC03767.1; -.
DR EMBL; AF106062; AAD45972.1; -.
DR EMBL; X86019; CAA60014.1; -.
DR MTM; 602357; -.
DR InterPro: IPR003124; WH2.
DR Pfam; PF02205; WH2; 1.
DR SMART; SM00246; WH2; 1.
KW Actin-binding; Repeat.
FT DOMAIN 2 13 POLY-PRO.
FT DOMAIN 64 96 GLY-RICH.
FT DOMAIN 241 244 POLY-SER.
FT DOMAIN 264 433 PRO-RICH.
FT REPEAT 352 361 XRSXPXPX MOTIF 1.
FT REPEAT 374 383 XRSXPXPX MOTIF 2.
FT REPEAT 410 419 XRSXPXPX MOTIF 3.
FT SITE 45 48 ACTIN BINDING.
FT VARIANT 495 495 G -> A.
FT /FTID=VAR_010295.
FT CONFLICT 303 309 PHRPHLR -> SSQAPP (IN REF. 3).
FT CONFLICT 360 360 P -> PV (IN REF. 2).
FT CONFLICT 487 503 GSNRRRGPPPLPIPR -> EYFCQGF (IN REF. 2).
SQ SEQUENCE 503 AA; 51489 MW; 43EB88674DD3BF1A CRC64;
```

alignment_scores:

```
Quality: 189.50 Length: 444
Ratio: 1.083 Gaps: 24
Percent Similarity: 39.414 Percent Identity: 27.477
```

alignment_block:

```

US-09-462-480-1/rev x WAIP_HUMAN
Align seg 1/1 to: WAIP_HUMAN from: 1 to: 503

1268 GCCGCATGACAACTCTCAGAGTGGCTCAACAGCTAATAACACAGAGAAA 1219
|||||
106 AlaGlyMetProLysLeuArgSerThrAlaAsnArgAspAsnAspSerG1 122
|||||
1218 GGGGAG.....ACGACGGAAGTTCGAACCTGCCTCCG 1187
|||||
122 yGlySerArgProProLeuLeuProProGlyGlyArgSerThrSerAlaL 139
|||||
1186 ATCCGGTGTTCGCTATTTCTACGGAACCTCGCGGTTCCTCCATCGCAACA 1137
|||||
139 ysPro...PheSerProProSerGlyProGlyArgPheProValProSer 154
|||||
1136 TCCCAAGTACGCTGCTTCGTCGACGCAATTCGCTGACCGCTTCGCTG 1087
|||||
155 ProGlyHisArg.....SerGlyProProGluProGlnArgAsnArg... 168
|||||
1086 ATCGTCCGCGCAGAGTTCCTGACGCGGTGTTCAGCTCGGTAGCGGTGC 1037
|||||
169 .....MetPro.....ProProA 173
|||||
1036 GTCCCATTTTCTCGTGGACACCTGTAG.....CTCCGAACGCTAC 993
|||||
173 rgProAspValGlySerLysProAspSerIleProProValProSer 189
|||||
992 CGCCCCAGCGCTCGCA...GCTTGGTCAGGACTGCTTCCCTCGTCA 946
|||||
190 ThrProArgProIleGlnSerSerLeuHisAsnArgGlySerProProVa 206
|||||
945 AGGAGGAATGAATGGAGCTGACATTTCCCTGATTGCTGCTGCGCGG. 897
|||||
206 lProgly.....GlyProArgGlnProSerP 215
|||||
896 .....CTCGATACCCGCGAATTCACCTGCTGCTGTCA 861
|||||
215 roGlyProThrProProPheProGlyAsnArgGlyThrAlaLeuGly 231
|||||
860 TGTTTTTCCTCGTTCTTTTCGATTAGCGGTTCAGAACCCATTCGG 811
|||||
232 GlyGlySerIleArgGlnSerProLeuSerSerSer..SerProPheSer 247
|||||
810 AGGACGCGCTGCT.....GCTGC 791
|||||
248 AsnArgProProLeuProProThrProSerArgAlaLeuAspAspLysPr 264
|||||
790 TCCTCGTGGCCCTCGAGTATTGGACGCGGCTGACGAATATTCGTGA 741
|||||
264 oProProProProProValGlyAsnArgPro.....SerIleH 278
|||||
740 GATCTCGTCGAGTCTCTGCTTCTGCTTATTCGCTCTTCTTGGAGCGCA 691
|||||
278 IsArgGluAlaValPro.....ProPro 285
|||||
690 CCACCG.....CGGCTGGCGGCGCTCCCGCGCGCGCGCCAC 650
|||||
286 ProProGlnAsnAsnLysProProValProSerThrProArgProSerAl 302
|||||
649 TGGCCCTCAACGAACCTCGCGTACCTCCACCTGCTGATCTGGGTTT 600
|||||
302 aProHisArgProHisLeuArgProProProProSerArgProGly.... 317
|||||
599 CAGGTGCGCGGAGATCCGCTGAAATTCCTCCCTGCTGCGGAGGTAG 550
|||||
318 .....ProProPro..... 320
|||||
549 CGGCATCGGTCTTCATCTCTGCGATGCTGGACTACTTCTCTCTTACCT 500
|||||
321 ..... 321
|||||
499 TCCTCGCCAAAATGTTGCAAGTCTTCCGG.CCCGGGTGGCGGGAAGTC 451

```

```

|||||
321 uProProSerSerGlyAsnAspGluThrProArgLeuProGlnArgA 338
|||||
450 TGTGTGTCATTACGGGAGCTCACCAGTCTCTCTTCGTCCTCCAGTCGCT 401
|||||
338 snLeuSerLeuSer.....SerSerThrProProLeuProSer..... 350
|||||
400 CGTCTCTTCTTCACGCTCTCGCGAGCGGTGCGCGCGCGACACGACCC 351
|||||
351 .....ProGlyArgSerGlyProLeuProProProPr 361
|||||
350 GGGCTGTGGAGCCCGCGGATTGGAAACCTGGCCCATCGCTCCCGGACC 301
|||||
361 oSer.....GluArgProProProValArgAspP 372
|||||
300 CACCGAGCGCGCCACCCGCTCACCGACGATCCGCGCAACCGCGCGCA 251
|||||
372 roProGlyArgSerGlyProLeuPro.....ProPro... 382
|||||
250 TCACCGAGGGGCAACCGCTTTTCGATCAGCTGAGACATCAGCGCGGTG 201
|||||
383 .....ProProValSerArgAsnGlySerThrSerArgAlaLe 395
|||||
200 CGGTCAACGACCCACCTGCGCCA.....GGTAGGACTCGCGCG 160
|||||
395 uProAlaThr..ProGlnLeuProSerArgSerGlyValAspSerProAr 411
|||||
159 CAGCAGCGCGCGCGCTGGGGCTGATCCACCGACGCGCGATGAT 110
|||||
411 gSerGlyProArgProProLeuProProAspArgPro..... 423
|||||
109 TCGACAGCGGACTGCTGTCGCGAGCAGCGCCATCTGCGCGGCTTCCTCTCG 60
|||||
424 .....Ser 424
|||||
59 GCTGGTTCGCGCGCGCGCTGCGCGCCAC 30
|||||
425 AlaGlyAlaProProProProSer 434
|||||
seq_name: SwissProt_40:ESA6_MYCLE

seq_documentation_block:
ID ESA6_MYCLE STANDARD; PRT; 95 AA.
AC Q50206; O33083;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 6 kDa early secretory antigenic target homolog (ESAT-6-like protein)
DE (L-ESAT).
GN ESAT6 OR ESX OR L45 OR ML0049 OR MLCB628.12C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Wieles B., Notenboom T., Naafs B., Offringa R., Ottenhoff T.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).

```

```
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -1- SIMILARITY: BELONGS TO THE ESA6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X90946; CAA2441.1; -
DR EMBL; Y14967; CAA75200.1; -
DR EMBL; AL583917; CAC29557.1; -
DR Leproma; ML0049; -
KW Complete proteome.
FT CONFLICT 55 55 R -> Q (IN REF. 1).
FT CONFLICT 90 90 M -> T (IN REF. 1).
SQ SEQUENCE 95 AA; 10465 MW; B1526F78CB2AB8A1 CRC64;

alignment_scores:
  Quality: 186.00      Length: 91
  Ratio: 2.447        Gaps: 0
  Percent Similarity: 83.516      Percent Identity: 36.264

alignment_block:
US-09-462-480-1 x ESA6_MYCLE

Align seg 1/1 to: ESA6_MYCLE from: 1 to: 95

869 CAGCAGTGGATTCGGGGGTATCGAGCGCGGCAAGCGCAATCCAGG 918
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 GlnAlaTrpHisPheProAlaLeuGlnGlyAlaValAsnGluLeuGlnG 19

919 AAATGTCAGTCCATTCCTCCCTTGTGAGGAGGGAAGCAGTCCCTGA 968
|::: :: ||| ||||| ||||| ||||| ||||| |||||
19 ySerGlnSerArgIleAspAlaLeuLeuGlnGlnCysGlnGluSerLeu 36

969 CCAAGCTCGCAGCGCTGGGGCGGTAGCGGTTCGGAGGCGTACAGGGT 1018
||||| ||||| ||||| ||||| ||||| ||||| |||||
36 hrLysLeuGlnSerSerTrpHisGlySerGlyAsnGluSerTyrSer 52

1019 GTCCAGCAAAATGGGACGCCAGCGCTACCGAGCTGAACACGCGCTGCA 1068
||||| ||||| ||||| ||||| ||||| ||||| |||||
53 ValGlnArgArgPheAsnGlnAsnThrGluGlyIleAsnHisAlaLeuG 69

1069 GAACCTGGCGGACGATCAGCGAGCGGTGTCAGCAATGGCTTCGACCG 1118
::: ||||| ||||| ||||| ||||| ||||| |||||
69 yAspLeuValGlnAlaIleAsnHisSerAlaGluThrMetGlnGlnThr 86

1119 AAGCAACGTCACCTGGGATGTC 1141
||| ||||| ||||| ||||| ||||| |||||
86 luAlaGlyValMetSerMetPhe 93

seq_name: SwissProt_40:CF10_MYCLE

seq_documentation_block:
ID CF10_MYCLE STANDARD; PRT; 99 AA.
AC Q33084;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 10 kDa culture filtrate antigen cfp10 homolog.
GN MLO050 OR MLCB628.13C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Eiglmeyer K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;
RL Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.
```

```
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=TN;
RA MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: BELONGS TO THE ESA6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y14967; CAA75210.1; -
DR EMBL; AL583917; CAC29558.1; -
DR Leproma; ML0050; -
KW Complete proteome.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 99 AA; 10833 MW; 859B484F7EFE5A8A CRC64;

alignment_scores:
  Quality: 182.00      Length: 99
  Ratio: 2.333        Gaps: 0
  Percent Similarity: 78.788      Percent Identity: 39.394

alignment_block:
US-09-462-480-1 x CF10_MYCLE

Align seg 1/1 to: CF10_MYCLE from: 1 to: 99

528 CGACGAGTGAAGACCGATCGCTACCTCGGCGGAGGAGGAGGAGTAATTT 577
||||| ||||| ||||| ||||| ||||| ||||| |||||
1 AlaGluMetIleThrGluAlaIleLeuThrGlnGlnAlaAlaGlnPh 17

578 CGACCGGATCTCGGCGACCTGAACACCGAGATCGACGAGTGGAGTCGA 627
|::: ||||| ||||| ||||| ||||| ||||| |||||
17 eAspGlnIleAlaSerGlyLeuSerGlnGluArgAsnPheValAspSerI 34

628 CGGCGAGTTCGTCAGGCGCAGTGGCGGCGGCGGCGGAGCGCGCC 677
::: ||||| ||||| ||||| ||||| ||||| |||||
34 leGlyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAlaAla 50

678 CAGGCGCGGTGGTGGCGCTTCCAAAGACGACCAATAAGCAGAGAAGCAGGA 727
::: ||||| ||||| ||||| ||||| ||||| |||||
51 LeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleArgG 67

728 ACTCGAGAGATCTCGACGAAATATTCGTACGCGCGGCGGTCCAAATCTCGA 777
|::: ||||| ||||| ||||| ||||| ||||| |||||
67 nLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnTyrThrL 84

778 GGGCGGACGAGGAGCAGCAGCGCGCTGCTCGCAAAATGGGCTC 824
::: ||||| ||||| ||||| ||||| ||||| |||||
84 yThrAspAspGluAlaAsnGlnLeuLeuSerSerLysMetAsnPhe 99

seq_name: SwissProt_40:VNUA_PVKA

seq_documentation_block:
ID VNUA_PVKA STANDARD; PRT; 1733 AA.
AC P33485;
DT 01-FEB-1994 (Rel. 28, Created)
```

DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Probable nuclear antigen.
 OS Pseudorabies virus (strain Kaplan) (PRV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxId=33703;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91021039; PubMed=2171211;
 RA Vilek C., Kozmik Z., Paces V., Schirm S., Schwyzler M.;
 RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
 oriented open reading frame: characterization of their promoter and
 enhancer regions.";
 RL Virology 179:365-377(1990).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M34651; AAA47471.1;
 DR PIR; B45344; B45344.
 FT DOMAIN 112 117 POLY-THR.
 FT DOMAIN 179 1733 GLY-RICH.
 FT DOMAIN 192 196 POLY-SER.
 FT DOMAIN 271 298 POLY-PRO.
 FT DOMAIN 304 308 POLY-ARG.
 FT DOMAIN 883 889 POLY-GLY.
 FT DOMAIN 1398 1405 POLY-GLY.
 SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8B8475BB5E2 CRC64;

alignment_scores:
 Quality: 172.50 Length: 504
 Ratio: 0.757 Gaps: 29
 Percent Similarity: 45.238 Percent Identity: 27.183

alignment_block:
 US-09-462-480-1 x VNUA_PRVKA ..
 Align seg 1/1 to: VNUA_PRVKA from: 1 to: 1733

```

31 GTGGCGCGCCGCGC.....GGCGG 50
||||| |||
840 ValGlyGlyGluGlyArgLeuGlyGlyProArgValGlyLeuAlaG 856

51 CAACCCAGCCGACGAGGAGCGGCGGCTGCTGGCAGCCAGTC 100
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
856 yArgSPAlaAlaGluAlaAlaValGlyArgGlyValLeuGlyHisGlyP 873

101 CG...CTGTGCAACATCCGCTG.....GCTGTGGATCAGCG 135
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
873 roGluArgAlaProGluProValLeuGlyGlyGlyGlyGlyGly 889

136 CCCAGCGCGCGCGCGCTG..... 156
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
890 GlnGlnArgGlySerGlyValArgSerGlyProGluSerGluGlyAlaAl 906

157 .....C 157

906 aLeuAlaProGlyProProValLeuPheValAlaValAlaVal 923

158 TCGCGCGCGGAG.....TCGCTACTT 177
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
923 alProAlaGluGlyArgAlaGlyGluProLeuValLeuAlaValPro 939

178 GCGCAGGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 227
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
940 GlyAlaAlaGly.....ProGlyArgAlaAlaLeuLeuLe 951

```

```

228 AAAGCCGTTGCCCTCGGTGATGATCGCGCGCTGTTCGCCGATCGTCGG 277
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
951 uAlaProLeuGlyArgTrpValArg...AlaGlyGlyGlyGlyAlaGlyV 967

278 TGACGGGTGGCGCGCTCCGCGGTGGTCCGGGAGCG..... 312
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
967 aAlaGlyGlyAlaGlyGluAlaGlyLeuGlyAlaGlyAlaGlyLeuGly 983

313 ATGGCCAGGTTTCGAATCGCGCGGTCCAC.CAGCCCGGTGTGGTCG 361
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
984 AlaGlyAlaGlyLeuGlyAlaGlyGlyAlaGlyGlyProGlyAlaGlyG 1000

362 CGCCGCGCACCTCGCGCGAGCGGTGAAGAAGACGA..... 398
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1000 uAlaGly.....GlyGlyAlaArgArgArgArgArgTrpA 1014

399 .....CGAGCAGCTGGCAGGAGA 419
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1014 sPAspGluAlaGlyLeuLeuGlyProGluArgGlyGlnAlaGlyArgGly 1030

420 GGACGACTGGTGAGCTCCGTAATGACACAGACTTCCC.....GGCCA 463
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1031 LeuArgGlyProGlyProArgGlyGlyLeuGlyGluProGlyArgGlyHI 1047

464 CCCGGCGCGAAGACTTCCCAACATTTTGGCGAGGAGGTAAA..... 506
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1047 sValGlyArg.....GlyGluGlyArgGlyValG 1058

507 .....GAG 509
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1058 LyProGlyGlyLeuAlaGlyAlaGlyProValHisAlaValAlaHisGln 1074

510 AGAAGTAGTCCAGCAGATGAGACCGGATGCCCTACCTCCCTCGG 559
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1075 ArgArgHisGlyAla...GlyAspGluGlyAspArgValArgGlyLeuPro 1090

560 GCAGGAGCGAGTAAATTTTCGAGCGGATCTCCGGCGACCTGAAACCCAGA 609
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1091 ProLeuGlyArgAlaGlyProGlyAsp...ArgValAlaGluArgGluG 1106

610 TCGACGAGTGGAGTCGACGCGAGTTCGTT.....GCAGGCGCAG 650
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1106 nArg.....GlyArgHisLeuLeuGluAlaGlyGlyProG 1118

651 TGGCGCGCGC.....GGCGGAGCGCGCGCGCGCGGT... 691
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1118 LuGlyGlyArgGlyAlaGlyGlyArgGlyGlnProGluArgAlaGlyGln 1134

692 ...GCGCTTCCAAGACGACCAATAGCAGAAGCAGGAACCTCGACAGAGA 738
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1135 GlnAlaLeuGluAspAlaAlaAlaGlyGlnAspAlaGlyValArgGln.. 1150

739 TCTCGAGCAATATTTCGTCAGCGCGCTCCAATACTCGAGGCGCGCAGAG 788
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1151 .LeuAlaGlyHisAlaAlaGlyLeuArg.....GlyGlyGluGlyG 1164

789 GAGCAGCAGCAGCGCTGCTCTCGCAATGGCTTCTGACCCGCTAATAC 838
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1164 LyAlaAspAlaGlyAlaGlyGlyLeuAspGlyArgLeuProGly..... 1178

839 GAAAGAAACGAGCAAAACATGACAGCAGCAGTGGATTTTCGCGGG 888
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1179 .....AlaG 1180

889 TATCAGGCGCGCGCAAGCGCAATCCAGGAAATGTCACGCTCCATTTCAT 938
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1180 yValArgGlyAlaAlaArgValGlyHisValGlyValGlyProAlaGluA 1197

939 CCCTCTTGGACGAGGAGCAGCTCCCTGACCAAGCTCGCAGCGCGCTGG 988
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1197 laLeuGlnAspGluGly.....LeuLeuGlyAlaIleValAlaAla 1211

```



```

1121 CTTCCGGTGAAGCCATTG...CTTGACCGCTTCGCCTG 1087
    ||| |||||:::||: |||||::|
825 LeuMetSerProValGlyArgIleLeuGluGluArgLeuGluPr 841
1086 ATCTCCCGGCCAGGTCTTGCACGCGGTGTTCAGCTCGGTACCGCTGGC 1037
    :|||||::|::|::|::|::|::|
841 oGlnAlaThrProGlyProThrGlnAlaAlaSerAla..... 853
1036 GTCCCATTTTCTCGACACCTGGTACCGCTCCGAAACCCTACCGCCCC 987
    |||||::|::|::|::|::|::|
854 .....PheGlyAlaGlyProThrGlnProHisProValValPro 867
986 AGCGCGCTGCGAGCTGGTCAGG..... 963
    |||||::|::|::|::|
868 SerProValArgProSerGlyGlyArgProGluGlyPheLeuGlyGlnAr 884
962 .....ACTGCTTCCCCTCGTCAAGGAGGGATGAATGACGTGCACATT 920
    ::||::|::|::|::|::|::|
884 gGlyArgAlaValArgProThrArgLeuSerLeuGluGlyLeuProSerL 901
919 TCCCTGGATTG...CGCTTGC CGCGCCCTCATACCGCGAATTCACCT 873
    :|||::|::|::|::|::|::|
901 euProSerMethHisGluThrProLeuProProGluProLysSerProGly 917
872 GCTGCTCTGTCATGTTTTGCTCCGCTTCTTTTCGTATTAGCGGGTCAGA 823
    ||||: |::|::|::|::|
918 GluTyrlleAsn.....IleAspPheGlyGl 926
822 AGCCCATTTGGCAGACAGCGCTGCTGCTGCTCTCTCGTCGGCCCTCG.. 775
    | ||| |||||::|::|::|::|::|
926 u..ProGlyAlaArgLeuSerProProAlaProProLeuLeuAlaSerAl 942
774 .....AGTATTG 768
    ||||: |::|::|
942 aAlaSerSerSerLeuLeuSerAlaSerSerProAlaLeuSerLeuG 959
767 GACGCGCGCCTGACGAATATTCGTGAGATCTCGT...CGAGTTCTCGCT 721
959 lySerGlyThrProGlyThrSerSerAspSerArgGlnArgSerProLeu 975
720 TCTGCTTATTGGCTGCTTCTTGGAGCCGACCA.....CCGCGGCTGG 677
    ||| ||| ||| |||||::|::|::|::|
976 SerAspTyrMetAsnLeuAspPheSerProLySserProLySprogl 932
676 CGCGCGCTGCC..... 667
    |||||
992 yAlaProSerGlyHisProValGlySerLeuAspGlyLeuLeuSerPro 1009
666 .....CGCGCGCGCGCGCCACTGCCCT 643
1009 luAlaSerSerProTyrProProLeuProArgProAlaSerPro 1025

```

menyus G.; May M.; Calle E.; Chambliss F.; D
RA
RT
TAF(II)135 potentiates transcription

RA Mengus G., May M., Carre L., Chambon P., Davidson I.:
RT "Human TAF(II)135 potentiates transcriptional activation by the AF-2s

314 A.....TCGCTCCGGACCCAGCGGAGCGGCCACCGCTCAC 277
| : : : : : | | | | | : : : : :
261 aAlaAlaProAlaAlaAlaProProProPro.....ProAlaP 275
276 CGACATCCGCAACAGCGCGGCATCACCGAGGGGCAACCGCTTTT 227
| : : : : : | | | | | : : : : :
275 foAlaThLeuAlaArgProGlyHisProAlaGlyProProThAla 291
| : : : : : | | | | | : : : : :
226 CGATCAGCT.....GAGCATCAGCGCGTGGTGGTCAACGACCA 186
| : : : : : | | | | | : : : : :
292 AlaProAlaValProProProAlaAlaAlaGlnAsnGlyGlySerAlaG1 308
185 CTGCGCCAGTAGGACTCCGCGGAGCAGCGCCCGCGCGCTGGG 136
| : : : : : | | | | | : : : : :
308 yAlaAlaProAlaProAlaProAlaAlaGlyGlyProAla.....G 322
135 GCCTGATCCACGACGCGGATGTTTCGACGCGGACTGTCGCGGACCA 86
| : : : : : | | | | | : : : : :
322 lyValSerGlyGlnProGlyProGlyAlaAlaAlaAlaAlaProAlaPro 338
85 GGCCCATCTGCGCGCTTCCTCGTGGTGGTGGTGGTGGTGGTGGTGG 36
| : : : : : | | | | | : : : : :
339 GlyValLysAlaGluSerProLysArgVal.....ValGlnAlaAla 352
35 CCCACTCGGTGACAAAG.....ACGTACCTGCTGCA 2
| : : : : : | | | | | : : : : :
352 aProAlaAlaGlnThrLeuAlaAlaSerGlyProAlaSer 366

seq_name: SwissProt_40:DRPL_HUMAN

seq_documentation_block:

ID DRPL_HUMAN STANDARD; PRT; 1185 AA.
AC P54259;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acrophin-1 (Dentatorubral-pallidolysian atrophy protein).
GN DRPLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Brain;
RX MEDLINE=95144175; PubMed=7842016;
RA Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,
RA Inoue T., Yamada M.;
RT "Structure and expression of the gene responsible for the triplet
RT repeat disorder, dentatorubral and pallidolysian atrophy (DRPLA).";
RL Nat. Genet. 8:177-182(1994).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96262314; PubMed=8965642;
RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,
RA Kidwai A.S., Ashworth R.G., Ross C.A.;
RT "DRPLA gene (atrophin-1) sequence and mRNA expression in human
RT brain.";
RL Brain Res. Mol. Brain Res. 36:219-226(1996).
[3]
RP SEQUENCE OF 470-725 FROM N.A.
RC TISSUE=Brain cortex;
RX MEDLINE=93315145; PubMed=8325628;
RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;
RT "Novel triplet repeat containing genes in human brain: cloning,
RT expression, and length polymorphisms.";
RL Genomics 16:572-579(1993).
CC -!- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,
CC OVARY, TESTIS AND PROSTATE. LOWER LEVELS ARE DETECTED IN THE
CC LIVER, THYMUS AND LEUCOCYTES.
CC -!- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC
CC (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO
CC ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN

CC CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE
CC DISEASE.
CC -!- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-
CC PALLIDOLYSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE
CC DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE
CC NUCLEUS, RUBRUM, GLOBUS PALLIDUS AND LUY'S BODY. CLINICAL FEATURES
CC ARE MYOCLONUS EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF
CC THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH
CC IN THE FOURTH.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC CC
CC EMBL; D31840; BAA06626.1; -;
CC EMBL; U23851; AAB50276.1; -;
CC EMBL; L10377; -; NOT_ANNOTATED_CDS.
CC HSSP; P00651; 1LRA.
CC MIM; 125370; -;
CC InterPro: IPR002951; Atrophin.
CC PRINTS; PR01222; ATROPHIN.
CC KW Triplet repeat expansion; Polymorphism.
CC FT DOMAIN 73 82 SER/GLU-RICH (MIXED CHARGE).
CC FT DOMAIN 302 305 POLY-PRO.
CC FT DOMAIN 376 382 POLY-SER.
CC FT DOMAIN 386 397 POLY-SER.
CC FT DOMAIN 442 447 POLY-PRO.
CC FT DOMAIN 479 483 POLY-HIS.
CC FT DOMAIN 484 497 POLY-GLN.
CC FT DOMAIN 504 507 POLY-PRO.
CC FT DOMAIN 564 574 POLY-SER.
CC FT DOMAIN 704 707 POLY-PRO.
CC FT DOMAIN 802 815 ARG/ALA-RICH (MIXED CHARGE).
CC FT DOMAIN 816 827 ARG/GLU-RICH (MIXED CHARGE).
CC FT DOMAIN 925 934 ARG/GLU-RICH (MIXED CHARGE).
CC FT CONFLICT 94 94 MISSING (IN REF. 2).
CC FT CONFLICT 333 333 Y -> H (IN REF. 2).
CC FT CONFLICT 339 339 M -> I (IN REF. 2).
CC FT CONFLICT 541 541 P -> T (IN REF. 3).
CC FT CONFLICT 1028 1028 G -> A (IN REF. 2).
CC SQ SEQUENCE 1185 AA; 124785 MW; 56C306267331C005 CRC64;

alignment_scores:

Quality: 165.00 Length: 536
Ratio: 0.721 Gaps: 35
Percent Similarity: 42.724 Percent Identity: 25.746

alignment_block:

US-09-462-480-1/rev x DRPL_HUMAN ..

Align seg 1/1 to: DRPL_HUMAN from: 1 to: 1185

1267 CGCGCATGACAACTCTCAGAGTGGCTCAACGTATAAACAACAGAGAAG 1218
| | | | | : : : : : | | | | : : : : :
255 ProProThrThrProIleSerValSerSer.....GlyAlaSerG1 269
1217 GCGGAGACCGGACGAAGTCCGACATCCCGCTCGGTTCGCTATTC 1168
| | | | | : : : : : | | | | : : : : :
269 yAlaProThrThrLysProProThrThrProValGlyGlyGlyAsnLeu. 285
1167 TACCGCAACTCGGCTTCCCTATCCGACATCCAGTCAGCTTCCCTTC 1118
| | | | | : : : : : | | | | : : : : :
286Pro.SerAlaProProAlaAsnPheProHisValThrProAs 300
1117 GGTCAAGACCATTCCTCCGACCGCTTCG.....TGATCG 1083
: : : : : | | | | : : : : :
300 nLeu.ProPro....ProProAlaLeuArgProLeuAsnAlaSerAla 315

```

1082 TCGCGCCAGTCTTCGAGCGCTGTTCAGCTCGGTAGCGTGGGCTC 1033
1083 ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
316 SerProGlyLeuGlyAla.....GlnPr 324
1032 CATTTTCTGGACACCGCTGGTACGCTCGGAACCGCTACCGC..... 990
1033 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
324 oLeu.....ProGlyHisLeuProSerProTyrAlaMetGlyG 337
989 .....CCGAGCGCGCTGCGAGCTTGTCTGAGGACT 960
990 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 lnglyMetGlyGlyLeuProGlyProGlyGlyProThrLeuAla 353
959 GCTTCCCTCGTCAAGGAGGAATGAATGACGTGACATTTCCT..... 915
960 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 ProSerProHisSer.....LeuProProAl 362
914 .....GGATTGCGCTTGGCGGGCTCGATACCGCGGAATTCACGTG 872
915 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
362 aSerSerAlaProAlaProMetArgPheProTyrSerSerS 379
871 CTGCTCTGTCATGTTTGTCTCGTCTTCTTCTGATTACGGGTGAGA. 823
824 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
379 erSerSerAlaAlaAlaSerSerSerSerSerSerSerSerSer 395
822 ...AGCCATTTCGAGGACAGCCCTGCTGCTGCTCTGCTGCGCCCTC 776
823 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
396 AlaSerProPhe.....ProAlaSerGlnAlaLeuProSerTyrProHi 410
775 G.....AGTATTGGACCGCGCCCTGACGAAAT 750
751 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
410 sSerPheProProThrSerLeuSerValSerAsnGlnProLysT 427
749 ATTCTGTCGAGATCTCGTGGAGTTCTGCTGCTTATTGCTGCTCTT 700
750 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
427 yrThrGlnProSer.....LeuProSerGlnAlaValTrp....SerGln 440
699 GGAAGCGCACCCCGCGCT...GGGCGG..... 673
700 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
441 GlyProProProProProProProProProProProProProPro 457
672 .....CGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
673 ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
457 aHisProGlyProPheProProSerThrGlyAlaGlnSerThraHisP 474
632 TGCGG..... 628
633 |||
474 roProValSerThrHisHisHisHisHisHisHisHisHisHis 490
627 .....TCGACTCCACCTGCTGCG 611
491 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 507
610 ATCTGGGTTT.....TCAGTCCGCGGAGATCCGCTCGAA 576
611 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
507 oGlyAlaPheProHisProLeuGluGlyGlySerSerHisHisAlaHisP 524
575 ATTAC...CTGCTCTCTCGCGGAGGTAGCGCATCGTCTTCTCTCTG 529
529 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
524 roTyrAlaMetSerProSerLeuGly.....SerLeu 534
528 CCATGCTGGAGTACTTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 488
529 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
535 ArgProTyrProProGlyProAlaHisLeuProProHisSerGlnVa 551
487 TGTGGCAAGTCTTCGCGCGCGGAGTGGCGGAGTCTGTGTGATPACG 438
551 lSerTyrSerGlnAlaGlyProAsnGly..... 560
437 GGAGCTCACCAGTCGCTCTTCTGCTCCAGTCTGCTCTGCTCTCTTCT 388
561 .....ProProVal.SerSerSerSerSerSerSerSerSerSer 575
387 ACGTCTCTGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 338

```

```

575 nGlySerTyrPro.....CysSerHisProSerProSerGlnG 588
589 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 CGCGGATTCGGAACCTGCGCCATCGCTCCCGGACCCACC..... 297
588 lyProGlnGlyAlaProTyrProPheProProValProThrValThr 604
297 ..... 297
605 SerSerAlaThrLeuSerThrValIleAlaThrValAlaSerSerProAl 621
296 .....GGAGCGGCGC 287
621 aGlyTyrLysThrAlaSerProGlyProProProTyrGly.LysArg 637
286 CACCGCTCACCGAGCATCCGCAA.....CAGCGCGCGCA...TCACC 246
638 AlaProSerProGlyAlaTyrLysThrAlaThrProGlyTyrLysPr 654
245 GAGGGGCAACCGCTTTCGATCAGTCAGATCAGCGCGCTGCGGGT 196
654 oGlySerProProSerPheArgThrGlyThrPro..ProGlyTyrArgG 670
195 CAACGACCCACCTGCG.....CCAGGTAGCGACTCCG 164
670 yThrSerProAlaGlyProGlyThrPheLysProGlySerProThrV 687
163 CGCGCAGCAGCGCGCGCGCGCTGGCGCTGATCCACCAGCCGCGGA 114
687 aGlyProGlyProLeuProProAlaGlyPro..... 697
113 TGGTTCACAGCGGACTGTCGCGAGCAGCGCCATCTGCGCGCTTCTCTC 64
698 .....SerGlyLeu...ProSerLeuProProProAlaAlaPr 710
63 GTCGCTGCTGTCGCGCGCGCGCTGGCGCGCGCGCGCGCGCGCGCG 15
710 oAlaSerGly.....ProProLeuSerAlaThrGlnIleLysGlnGlu 724
seq_name: SwissProt_40:CAFF_RIFPA
seq_documentation_block:
ID CAFF_RIFPA STANDARD; PRT; 1027 AA.
AC P30754;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Fibril-forming collagen alpha chain.
OS Riffia pachyptila (Tube worm).
OC Eukaryota; Metazoa; Vestimentifera; Axonobranchia; Riffiida;
OC Riffidae; Riffia.
OX NCBI_TaxID=6426;
RN [1]
RP SEQUENCE.
RX MEDLINE=93130909; PubMed=1483468;
RA Mann K., Gailli F., Timpl R.;
RT "Amino-acid sequence and cell-adhesion activity of a fibril-forming
RT collagen from the tube worm Riffia pachyptila living at deep sea
RT hydrothermal vents.";
RL Eur. J. Biochem. 210:839-847(1992).
RN [2]
RP SEQUENCE OF 8-45; 525-618 AND 810-882.
RC TISSUE=Cuticle;
RX MEDLINE=92015209; PubMed=1920405;
RA Gailli F., Wiedemann H., Mann K., Kuhn K., Timpl R., Engel J.;
RT "Molecular characterization of cuticle and interstitial collagens
RT from worms collected at deep sea hydrothermal vents.";
RL J. Mol. Biol. 221:209-223(1991).
CC -!- FUNCTION: FIBRIL-FORMING COLLAGEN.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- PTM: ALL THE "X" ARE PROBABLY GLYCOSYLATED HYDROXYLYSINES.
DR PIR; S22915; S22915.
DR InterPro; IPR000087; Collagen.

```



```

1083 CGATCAGCGAAGCCGGTCAGGC 1104
      I          :      |||
      762 pGlyThr**GlyGluArgGly 769

seq_name: SwissProt_40:CA21_BOVIN

seq_documentation_block:
ID   CA21_BOVIN          STANFORD;          PRT; 1364 AA.
AC   P02465; O62649;
DT   21-JUL-1986 (Rel. 01, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen alpha 2(I) chain precursor.
GN   COL1A2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=98290219; PubMed=9628255;
RA   Shirai T., Hattori S., Sakaguchi M., Inouye S., Kimura A., Ebihara T.
RA   Irie S., Nagai Y., Hori H.;
RT   "The complete cDNA coding sequence for the bovine proalpha2(I) chain
RT   of type I procollagen."
RL   Matrix Biol. 17:85-88(1998).
RN   [2]
RP   SEQUENCE OF 80-98.
RC   TISSUE=Skin;
RX   MEDLINE=75036115; PubMed=4609475;
RA   Fietzek P.P., Breitkreutz D., Kuehn K.;
RT   "Amino acid sequence of the amino-terminal region of calf skin
RT   collagen."
RL   Biochim. Biophys. Acta 365:305-310(1974).
RN   [3]
RP   SEQUENCE OF 95-415, AND REVISION.
RC   TISSUE=Skin;
RX   MEDLINE=76091874; PubMed=173531;
RA   Fietzek P.P., Rexrodt F.W.;
RT   "The covalent structure of collagen. The amino-acid sequence of
RT   alpha2-CB4 from calf-skin collagen."
RL   Eur. J. Biochem. 59:113-118(1975).
RN   [4]
RP   SEQUENCE OF 416-445.
RC   TISSUE=Skin;
RX   MEDLINE=75008198; PubMed=4412529;
RA   Fietzek P.P., Furthmayr H., Kuehn K.;
RT   "Comparative sequence studies on alpha2-CB2 from calf, human, rabbit
RT   and pig-skin collagen."
RL   Eur. J. Biochem. 47:257-261(1974).
RN   [5]
RP   SEQUENCE OF 446-481.
RC   TISSUE=Skin;
RX   MEDLINE=75059250; PubMed=4435743;
RA   Fietzek P.P., Kuehn K.;
RT   "The covalent structure of collagen: amino acid sequence of the N-
RT   terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.S
RT   from calf skin collagen."
RL   Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).
CC   -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC   (FIBRILLAR FORMING COLLAGEN).
CC   -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC   BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC   HYDROXYAPATITE.
CC   -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at Heidelberg.

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AB008683; BAA25171.1; -.
 DR PIR; A02866; CGB025.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 17.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Collagen; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT PROPEP 25 79 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 80 1100 COLLAGEN ALPHA 2(I) CHAIN.
 FT PROPEP 1101 1364 CARBOXYL-TERMINAL PROPEPTIDE.
 FT MOD_RES 80 80 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 84 84 CONVERTED TO AN ALDEHYDE GROUP THAT IS
 FT INVOLVED IN CROSS-LINKING (PROBABLE).
 FT MOD_RES 175 175 HYDROXYLATION (PROBABLE).
 FT MOD_RES 196 196 HYDROXYLATION (PROBABLE).
 FT MOD_RES 262 262 HYDROXYLATION (PROBABLE).
 FT MOD_RES 307 307 HYDROXYLATION (PROBABLE).
 FT MOD_RES 352 352 HYDROXYLATION (PROBABLE).
 FT CONFLICT 157 157 V -> P (IN REF. 3).
 FT CONFLICT 187 187 K -> T (IN REF. 3).
 FT CONFLICT 211 211 T -> K (IN REF. 3).
 FT CONFLICT 298 300 PGA -> AGP (IN REF. 3).
 FT CONFLICT 423 424 AT -> TA (IN REF. 4).
 SQ SEQUENCE 1364 AA; 129064 MW; 5593F4D6B9ED119A CRC64;

alignment_scores:

Quality: 164.50 Length: 396
 Ratio: 0.848 Gaps: 23
 Percent Similarity: 48.990 Percent Identity: 27.273

alignment_block:

US-09-462-480-1 x CA21_BOVIN ..
 Align seg 1/1 to: CA21_BOVIN from: 1 to: 1364
 31 GTGGCGGCGACCGCGCGGCAACCCAGCGACGAGGAAGCGCGCAGAT 80
 268 ValGlyAsnProGlyProAlaGlyProAlaGlyProArg...GlyGluva 283
 81 GGGCGCTGCTGGCACCAGTCCGCTGTCTCAACCAATCCGCTGGTGGAT 130
 283 lGlyLeuProGly.....LeuSerGlyProV 292
 131 CAGCGCCCGCGCGCGCGGCGCTGTGCGCGGAGTCTACCTGGC 180
 292 alGlyProProGlyAsnProGly.....AlaAsnGlyLeuProGly 305
 181 GCAGGTGGTGTGTTGACCCGACGCGCGCTGATGCTCAGCTGATCGAAAA 230
 306 AlaLysGlyAla...AlaGlyLeuProGlyValAlaGlyAlaProGlyLe 321
 231 GCGGTTGCCCTCGGTGTGATGCGCGCGGTGTGCGGATCGTCGGTGA 280
 321 uPro...GlyProArgGlyLeProGlyProValGlyAlaAlaGlyAlat 337
 281 CGGTGGCGCGCTCCCGTGGGT...CCGGAGCGGATGGCCAGGGTTCG 327
 337 hrGlyAlaArgGlyLeuValGlyLeuProGlyProAlaGlySerLysGly 353
 328 CAATCCGGCGGCTCCACGACGCGGCTGTGTGCG...CCGGCACCCT 374
 354 GluSerGlyAsnLysGlyGluProGlyAlaValGlyGlnProGlyProPr 370

375 CGCGCAGGAGCGTGAAGAAGACGACGAGGACGACTGGGACGAAGAGGACG 424
 370 oGlyProSerGlyGlu.....Glu.GlyLysArgGlySer 381
 425 ACTGGTGAAGTCCCGTAATGACACACAGACTCCCGCGCCACCGCGCCGGA 474
 382 ThrGlyGluIle.....GlyProAlaGlyProProGlyProPr 394
 475 AGACTTGCACACATTTTGGCGGAGGAAGTAAAGAGAGAAAGTAGTCCAG. 523
 394 oGlyLeu.....ArgGlyAsnProGlySerArgGlyLeuProG 407
 524CATGGCAGAGATGAAGACCCGATGCCGCG 550
 407 lAlaAspGlyArgAlaGlyValMetGlyProAlaGlySerArgGlyAla 423
 551 TACCTCGGCGGAGGAGGAGGATTTTCGAGCGGATCTCCGCGGACCTGA 600
 424 ThrGlyProAlaGlyValArgGlyProAsnGlyAspSerGlyArgProGl 440
 601 AAAC.....CCAGATCGACCGAGTGGAG 623
 440 yGluProGlyLeuMetGlyProArgGlyPheProGlySerProGlyAsnI 457
 624 TCGAGCGGAGGTTCTGTCAGGGCGAGTGGCGCGCGCGCGGAGCGGC 673
 457 leGlyProAlaGlyLysGlyGlyProValGlyLeuProGlyIleAspGly 473
 674 CGCCCGGCGCGGTGGTGGCTTCCAGAGACGACCAATAGCAGAGACG 723
 474 ArgProGlyProIleGly.....ProAl 481
 724 AGGAACTCGAGCAGATCTCGAGCAATATTCGTCAGGCGCGGTCCAA... 770
 481 aGlyAlaArgGlyGlu.ProGlyAsnIleGlyPheProGlyProLysGly 497
 771TACTCGAGGCGGACGAGGACGAGCAGCAGCGGCGCTGC 808
 498 ProSerGlyAspProGlyLysAlaGlyGlyLysGlyHisAlaGlyLeuAl 514
 809 CTCGCAATGGCTTCGACCGCTAATACGAAAGAACGAGCAACAAA 858
 514 aGlyAlaArgGlyAla...ProGlyProAspGlyAsnAsnGlyAlaGlnG 530
 859 CATGACAGACGACGAGTGGATTTTCGCGGGTATCGAGCGCGCGCAAGCG 908
 530 lyProProGlyLeuGlnGlyVal.GlnGly.....GlyLysGlyGluGl 544
 909 CAATCCA...GGGAAATGTACGTCCTCATTCCTCTCTTTCAGCAGGGG 955
 544 nGlyProAlaGlyProProGlyPheGlnGlyLeuProGlyProAlaGlyT 561
 956 AAGCA.....GTCCCTGACCAAGCT 975
 561 hrAlaGlyGluAlaGlyLysProGlyGluArgGlyIleProGlyGluPhe 577
 976 CGCAGCGCGCTGGGCGGTAGCGGTTCGAGGCGGTACACAGGCTGTCCAGC 1025
 578GlyLeuProGlyProAlaGly 584
 1026 AAAAATGGAGCGCCACGCTACCGAGCTGAACACGCGCTGCAGAACCT. 1074
 584 yAlaArgGlyGluArgGlyProProGlyGlyLysGlyAlaAlaGlyProt 601
 1075 ..GGCGGCGGAGTACCGAGCGCGGTGAGC 1104
 601 hrGlyProIleGlySerArgGlyProSerGly 611
 seq_name: SwissProt_40.MY15_HUMAN
 seq_documentation_block:
 ID MY15_HUMAN STANDARD; PRT; 3530 AA.

DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00660; BAND_41_1; FALSE_NEG.
DR PROSITE; PS00661; BAND_41_2; FALSE_NEG.
DR PROSITE; PS00507; BAND_41_3; 1.
DR PROSITE; PS00096; IQ; 3.
DR PROSITE; PS00002; SH3; FALSE_NEG.
KW Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW Calmodulin-binding; Disease mutation; Deafness.
FT DOMAIN 1 1887 HEAD OR MOTOR DOMAIN.
FT DOMAIN 1888 2029 NECK OR REGULATORY DOMAIN.
FT DOMAIN 2030 3530 TAIL.
FT DOMAIN 1323 1350 COILED COIL (POTENTIAL).
FT DOMAIN 1792 1799 ACTIN-BINDING (POTENTIAL).
FT DOMAIN 1902 1924 IQ 1.
FT DOMAIN 1925 1954 IQ 2.
FT DOMAIN 1955 1976 IQ 3.
FT DOMAIN 2867 2953 SH3.
FT DOMAIN 3206 3443 BAND 4.1-LIKE.
FT NP_BIND 1315 1322 ATP (POTENTIAL).
FT VARIANT 2111 2111 N -> Y (IN DFNB3; FAMILY FROM BENGKALA).
FT VARIAT 2113 2113 /FTID=VAR_010303.
FT SEQUENCE 3530 AA; 395171 MW; 3D103923D4BCBE4A CRC64;

alignment_scores:
Quality: 163.50 Length: 481
Ratio: 0.790 Gaps: 27
Percent Similarity: 43.035 Percent Identity: 26.611

alignment_block:
US-09-462-480-1/rev x MY15_HUMAN ..
Align seg 1/1 to: MY15_HUMAN from: 1 to: 3530

1262 ATGACAACCTCTCAGATCGCGTCAAAAGTATAAACACGAGAAGGGCGGA 1213
:::| | | | | ::::| | | | |
535 LeuThrProArgGlnArgAsnLeuGlnArgAlaLeuSerAlaPheGlyAl 551
1212 GACCGACGGAGGTCAAATCCGCCGAT.....CCCGTGTTTCGGCT 1172
| | | : : : | | | : : : | | | | | | | | |
551 aHisArgGlyLeuGlyPheGlyProGluPheGlyArgProValProArgP 568
1171 ATTCTACGC GAATCGCGCTGCCTATCGCAACATCCACGTCAGCTTGC 1122
: : : | | | | | | | | | : : : | | | : : :
568 roAlaThrSerLeuAlaArgPheLeuLysLysThrLeuSer..... 581
1121 CTTCGGTCGAAGCATTCGCTGACCGGCTCGCTGATCGTCC..... 1080
:::::| | | | | : : : | | | | | | | | |
582 ...GluLysLysProIleAla...ArgLeuArgGlySerGlnLysThrAr 596
1079 .GCCCCAGTTTCGACGCGGTTGTCACGTCGTAGCCGTCGGCGCTCCA 1031
| | | | | | | | | : : : | | | : : :
596 gAlacGlyProAlaValArgGluAlaAlaTrpLysArgPheGlyTrpL 613
1030 TTTTGTGCTGGA.....CACCCCTGGTAGCGCTCCGAACCGCTAC 993
: : : | | | | | | | | | | | | | | | | | |
613 ySLeuAlaGlyMetAspProGluLysProGlyThrProIleValLeuArg 629
992 CGCCCCAGCGCGTGGAGCTGGTCAGGACTGGTCC....CCTCGTCT 947
| | | : : | | | | | | | | | | | | | | | | | |
630 ArgAlaGlnProArgAlaArgSerSerAsnAspAlaArgArgProAla 646
946 AAGAGGGAAATGAATGGACGTGACATTCCCTGGATTGCGCTCCCGCGG 897
: : : : : : : : : | | | | | | | | | : : :
646 aProGlnProAlaProArgThrLeuSerHisTrpSerAlaLeuLeuSerP 663
896 CCTCGATACC CGCGGAATCCACTGCTGCTGTCTCATGTTTTTTCCTCGT 847
: : : | | | : : : | | | : : : | | | : : :
663 roProValProProArg..... 668
846 TTCTTTTCGTATTAGCGGGTCAAGGCCATTTGGCAGGACAGCGCTCGC 797

RC TISSUE=Placenta;
 RX MEDLINE-94375010; PubMed=8088784;
 RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W.,
 RA Uitto J., Greenspan D.S.;
 RT "Structural organization of the human type VII collagen gene (COL7A1),
 RT composed of more exons than any previously characterized gene.";
 RL Genomics 21:169-179(1994).
 RN [8]
 RP REVIEW ON DEB VARIANTS.
 RX MEDLINE-98041696; PubMed=9375848;
 RA Jaervikallio A., Pulkkinen L., Uitto J.;
 RA "Molecular basis of dystrophic epidermolysis bullosa: mutations in
 RT the type VII collagen gene (COL7A1).";
 RL Hum. Mutat. 10:338-347(1997).
 RN [9]
 RP VARIANT RDEB LYS-2798.
 RX MEDLINE-93291877; PubMed=8513326;
 RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y.,
 RA Lin A.N., Dietz H.C., Hovnanian A., Uitto J.;
 RT "A missense mutation in type VII collagen in two affected siblings
 RT with recessive dystrophic epidermolysis bullosa.";
 RL Nat. Genet. 4:62-66(1993).
 RN [10]
 RP VARIANT DDEB SER-2040.
 RX MEDLINE-94224777; PubMed=8170945;
 RA Christiano A.M., Ryyanen M., Uitto J.;
 RT "Dominant dystrophic epidermolysis bullosa: identification of a
 RT Gly-->Ser substitution in the triple-helical domain of type VII
 RT collagen.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994).
 RN [11]
 RP VARIANT PEB-DDEB CYS-2623.
 RX MEDLINE-96081220; PubMed=8541842;
 RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Uitto J.;
 RA "Pretibial epidermolysis bullosa: genetic linkage to COL7A1 and
 RT identification of a glycine-to-cysteine substitution in the triple-
 RT helical domain of type VII collagen.";
 RL Hum. Mol. Genet. 4:1579-1583(1995).
 RN [12]
 RP VARIANT DDEB ARG-2043.
 RX MEDLINE-95164985; PubMed=7861014;
 RA Christiano A.M., Moricone A., Paradisi M., Angelo C., Mazzanti C.,
 RA Cavalieri R., Uitto J.;
 RT "A glycine-to-arginine substitution in the triple-helical domain of
 RT type VII collagen in a family with dominant dystrophic epidermolysis
 RT bullosa.";
 RL J. Invest. Dermatol. 104:438-440(1995).
 RN [13]
 RP VARIANTS RDEB AND DDEB.
 RX MEDLINE-96220218; PubMed=8644729;
 RA Christiano A.M., McGrath J.A., Tan K.C., Uitto J.;
 RT "Glycine substitutions in the triple-helical region of type VII
 RT collagen result in a spectrum of dystrophic epidermolysis bullosa
 RT phenotypes and patterns of inheritance.";
 RL Am. J. Hum. Genet. 58:671-681(1996).
 RN [14]
 RP VARIANT RDEB ARG-2575.
 RX MEDLINE-96154068; PubMed=8592061;
 RA Shimizu H., McGrath J.A., Christiano A.M., Nishikawa T., Uitto J.;
 RT "Molecular basis of recessive dystrophic epidermolysis bullosa:
 RT genotype/phenotype correlation in a case of moderate clinical
 RT severity.";
 RL J. Invest. Dermatol. 106:119-124(1996).
 RN [15]
 RP VARIANT RDEB ARG-1782.
 RX MEDLINE-96183562; PubMed=8618018;
 RA Christiano A.M., McGrath J.A., Uitto J.;
 RT "Influence of the second COL7A1 mutation in determining the
 RT phenotypic severity of recessive dystrophic epidermolysis bullosa.";
 RL J. Invest. Dermatol. 106:766-770(1996).
 RN [16]
 RP VARIANT RDEB ASP-2073.
 RX MEDLINE-96310789; PubMed=8757758;
 RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,
 RA Uitto J., Pope F.M., Eady R.A.J.;
 RT "Clinicopathological correlations of compound heterozygous COL7A1
 RT mutations in recessive dystrophic epidermolysis bullosa.";
 RL J. Invest. Dermatol. 107:171-177(1996).
 RN [17]
 RP VARIANTS RDEB W-1982; G-2008; A-2025; E-2049; G-2063; W-2063 AND
 RP R-2575.
 RX MEDLINE-97465605; PubMed=9326325;
 RA Hovnanian A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,
 RA Fratang S., Christiano A.M., Uitto J., Lathrop M., Barrandon Y.,
 RA de Prost Y.;
 RT "Characterization of 18 new mutations in COL7A1 in recessive
 RT dystrophic epidermolysis bullosa provides evidence for distinct
 RT molecular mechanisms underlying defective anchoring fibril
 RT formation.";
 RL Am. J. Hum. Genet. 61:599-610(1997).
 RN [18]
 RP VARIANT RDEB ARG-1652.
 RX MEDLINE-98106792; PubMed=9444387;
 RA Cserhalmi-Friedman P.B., Karpati S., Horvath A., Christiano A.M.;
 RT "Identification of a glycine substitution and a splice site mutation
 RT in the type VII collagen gene in a proband with mitis recessive
 RT dystrophic epidermolysis bullosa.";
 RL Arch. Dermatol. Res. 289:640-645(1997).
 RN [19]
 RP VARIANTS DEB ARG-2009 AND ARG-2043.
 RX MEDLINE-97358588; PubMed=9215684;
 RA Winberg J.-O., Hammami-Hausali N., Nilsson O., Anton-Lamprecht I.,
 RA Naylor S.L., Kerbacher K., Zimmermann M., Krajci P.,
 RA Gedde-Dahl T. Jr., Bruckner-Tuderman L.;
 RT "Modulation of disease severity of dystrophic epidermolysis bullosa by
 RT a splice site mutation in combination with a missense mutation in the
 RT COL7A1 gene.";
 RL Hum. Mol. Genet. 6:1125-1135(1997).
 RN [20]
 RP VARIANTS DDEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.
 RX MEDLINE-98334662; PubMed=9668111;
 RA Hammami-Hausali N., Schumann H., Raghunath M., Kilgus O., Luethi U.,
 RA Luger T., Bruckner-Tuderman L.;
 RT "Some, but not all, glycine substitution mutations in COL7A1 result in
 RT intracellular accumulation of collagen VII, loss of anchoring
 RT fibrils, and skin blistering.";
 RL J. Biol. Chem. 273:19228-19234(1998).
 RN [21]
 RP VARIANTS DEB CYS-2008; ARG-2207 AND SER-2775.
 RX MEDLINE-98410969; PubMed=9740253;
 RA Kon A., Pulkkinen L., Ishida-Yamamoto A., Hashimoto I., Uitto J.;
 RT "Novel COL7A1 mutations in dystrophic forms of epidermolysis
 RT bullosa.";
 RL J. Invest. Dermatol. 111:534-537(1998).
 RN [22]
 RP VARIANT RDEB ARG-1347.
 RX MEDLINE-99019477; PubMed=9804332;
 RA Terracina M., Posteraro P., Schubert M., Sonego G., Atzori F.,
 RA Zamburano G., Bruckner-Tuderman L., Castiglia D.;
 RT "Compound heterozygosity for a recessive glycine substitution and a
 RT splice site mutation in the COL7A1 gene causes an unusually mild form
 RT of localized recessive dystrophic epidermolysis bullosa.";
 RL J. Invest. Dermatol. 111:744-750(1998).
 RN [23]
 RP VARIANTS DEB TRP-2034; VAL-2040; ARG-2043; ARG-2064 AND ASP-2713.
 RX MEDLINE-99072663; PubMed=9856843;
 RA alignment_scores:
 Quality: 163.00 Length: 433
 Ratio: 0.827 Gaps: 23
 Percent Similarity: 45.497 Percent Identity: 25.635
 alignment_block:
 US-09-462-480-1 x CA17_HUMAN ..
 Align seg 1/1 to: CA17_HUMAN from: 1 to: 2944

```
27  CCAGGTGGCGCGCACCGCGCGCGCAACCCAGCCGACGAGGAGCGCGC 76
    ||||| :|: ||| |||:|: ||| :|:|:|:
1464 ProGlyGluGlnGlyProArgGlyProGlyAlaIleGlyProLysGI 1480
    77 AGAT.....GGGCTGCTCGGCACGAGTCCGCTCGAACCAT 114
    :||| ||||| ||||| :|: :|:|:|:|:
1480 yAspArgGlyPheProGlyProLeuGlyGluAlaGlyGluLysGly 1496
    115 CCGCTGGCTGTGATCAGCGCCGCGCGCGCGCGCGCTGCTCGCGC 164
    ||| |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1497 .....ArgGlyProProGlyProAlaGlySerArgGlyLeuProGlyVa 1511
    165 GGAGTGCCTACTGGCGAGGTGGTCTGACCCGCGCGCGCTGATGT 214
    : :|: ||||| ||||| :|: :|:|:|:|:|:
1511 lAlaGlyArgProGlyAlaLysGlyProGluGlyProProGlyProThrG 1528
    215 CTCAGCTGATCGAAAG.....CGGTTGCCCTCGGTG 249
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1528 lYArgGlnGlyGluLysGlyGluProGlyArgProGlyAspProAlaVal 1544
    250 ATGCCGGCGGTGTCGGA....TCGTCGTGACGGTGGCGCGCTCC 296
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1545 ValGlyProAlaValAlaGlyProLysGlyGluLysGlyAspValGlyPr 1561
    297 GGTGGGTCCG...GGAGCGATGGGCCAGGTTCCGAATCCGCGGCTCCA 343
    |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1561 oAlaGlyProArgGlyAlaThrGlyValGlnGlyGluArgGly..... 1575
    344 CCAGCCGGGTGTGGTCCGCGCGCGCGCTGCGCAGGAGCGTGAAGAA 393
    ||||| ||||| ||||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1576 ..ProProGlyLeuValLeuProGlyAspProGlyProLysGlyAspPro 1591
    394 GACGACGAGG.....ACGACTGGGACGAGGAGGAGTGGT 431
    ||| ||| ||| ||||| |||||
1592 GlyAsp.ArgGlyProIleGlyLeuThrGlyArgAlaGlyProProGlyA 1608
    432 AGCTCCGCTAATGACACACTTCCCGGCGCACCCGCGCGGAGACTTG 481
    :|: ||| ||||| ||||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1608 sp.....SerGlyProProGlyGluLysGlyAsp 1617
    482 CCACATTTTGGCAGGAGTAAGAG.....AAGTAGTCCAGCA 525
    ||| ||||| ||||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1618 Pro.....GlyArgProGlyProProGlyProValGlyProAr 1630
    526 TGGCAGAGATGAAGACCGATCCGCTACCTCGGCGAGGAGGTAAT 575
    :||| ||||| |||||
1630 gGlyArgAsp.....GlyGluValGlyG 1638
    576 TTCGAGCGGATCTCCGCGCACTGAAACCCAGATCGACAGGT..... 619
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1638 LuLysGlyAspGluGlyProProGlyAspProGlyLeuProGlyLysAla 1654
    620 .....GGAGTCGCGGAGGTTCTGTCGAGGCGCAGT 651
    ||||| ||||| ||||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1655 GlyGluArgGlyLeuArgGlyAlaProGly.....ValArgGlyProVa 1669
    652 GCGCGCGCGCGCGGCGCGCGCGCGCGCGCGGTGGTGGCTTCCAA 701
    |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1669 lGlyGluLysGlyAspGlnGlyAspProGlyGluAspGly.....A 1683
    702 GAAGCAGCCAATAGCAGAGCAGGAACTCGACGAGATCT..... 741
    ||||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1683 rgAsnGlySerProGlySerSerGlyProLysGlyAspArgGlyGluPro 1699
    742 .....CGACGAATATTCTGTCAGGCGCGGTCCAA 770
    ||||| ||||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1700 GlyProProGlyProProGlyArgLeuValAspThrGlyProGlyAlaAr 1716
    771 TACTCGAGGCGCGACGAGGAGCAGCAGCGCTGCTCGCAAAATGGG 820
    : :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1716 gGluLysGlyGluProGlyAspArgGlyGlnGluGlyProArgGlyProL 1733
```

OM of: US-09-462-480-1 to: SPTRMBL_19:* out_format : pfs
Date: Jul 22, 2002 1:39 AM
About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+n2p_model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09462480/runat_18072002_164419_19544/app_query.fasta_1.2850
-DB=SPTRMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462480 -CGN1_133 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-462-480-1
Query length: 1277
Database: SPTRMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 143.350000

score_list:	Sequence	Strd Orig	zScore	Escore	Len	Documentation
	sp_bacteriap:069738	+ 713.00	869.06	1.3e-40	368	! O69738 mycobacterium tuberculosis
	sp_plant:Q9ZW08	- 210.50	246.19	2.6e-06	891	! Q9ZW08 arabidopsis thaliana (md
	sp_plant:Q9YFB2	+ 203.50	246.16	2.6e-06	894	! Q9YFB2 arabidopsis thaliana (md
	sp_human:Q9UQ39	+ 210.50	236.98	7.8e-06	956	! Q9UQ39 homo sapiens (human). r
	sp_human:Q9UQ40	+ 203.50	234.42	8.2e-06	1262	! Q9UQ40 homo sapiens (human). s
	sp_human:Q9UQ48	+ 203.50	228.91	9.1e-06	2296	! Q9UQ48 homo sapiens (human). s
	sp_human:Q9UQ35	+ 203.50	227.24	9.4e-06	2752	! Q9UQ35 homo sapiens (human). r
	sp_bacteriap:Q33085	+ 198.00	240.87	1.5e-05	302	! Q33085 mycobacterium leprae. hy
	sp_human:Q15038	+ 195.50	221.45	3.0e-05	1783	! Q15038 homo sapiens (human). k
	sp_human:Q60382	+ 195.50	221.41	3.1e-05	1791	! Q60382 homo sapiens (human). k
	sp_human:Q9H5F9	- 193.00	226.95	3.8e-05	704	! Q9H5F9 homo sapiens (human). c
	sp_bacteria:Q9XPD2	- 193.00	226.21	3.9e-05	763	! Q9XPD2 mycobacterium tuberculosis
	sp_rodent:Q70495	- 192.00	223.50	4.7e-05	897	! Q70495 mus musculus (mouse). pl
	sp_human:Q9UQ36	+ 189.00	216.59	8.0e-05	1275	! Q9UQ36 homo sapiens (human). r
	sp_human:Q9Y2J6	- 185.00	216.32	0.0001	772	! Q9Y2J6 homo sapiens (human). k
	sp_human:Q9ULT1	- 184.50	212.25	0.0002	1124	! Q9ULT1 homo sapiens (human). k
	sp_virus:Q9Q5K9	- 182.50	215.46	0.0002	608	! Q9Q5K9 herpesvirus papio. ntr.
	sp_human:Q60585	- 180.50	210.26	0.0003	820	! Q60585 homo sapiens (human). se
	sp_bacteria:Q9EVZ2	- 180.00	220.55	0.0002	251	! Q9EVZ2 paracoccus pantotrophus.
	sp_human:Q9UQ36	- 179.50	204.97	0.0004	1275	! Q9UQ36 homo sapiens (human). r
	sp_human:Q15038	- 179.50	201.88	0.0004	1783	! Q15038 homo sapiens (human). k
	sp_human:Q60382	- 179.50	201.84	0.0004	1791	! Q60382 homo sapiens (human). k
	sp_human:Q9UQ35	- 179.50	197.88	0.0004	2752	! Q9UQ35 homo sapiens (human). r
	sp_virus:Q41972	+ 179.00	209.80	0.0003	706	! Q41972 murid herpesvirus 4. hyp
	sp_virus:Q41973	+ 179.00	209.53	0.0003	727	! Q41973 murid herpesvirus 4. hyp
	sp_virus:Q69340	+ 178.50	199.79	0.0004	1958	! Q69340 pseudorabies virus. ori
	sp_virus:Q41971	+ 177.50	207.12	0.0004	774	! Q41971 murid herpesvirus 4. hyp
	sp_virtebrate:Q90YB5	- 177.00	206.39	0.0005	784	! Q90YB5 gallus gallus (chicken). c
	sp_rodent:Q35126	- 176.50	202.05	0.0006	1175	! Q35126 mus musculus (mouse). c
	sp_rodent:P70200	- 176.50	202.05	0.0006	1175	! P70200 mus musculus (mouse). c
	sp_plant:Q9G397	+ 176.00	207.04	0.0005	640	! Q9G397 chilamydomonas reinhardtii
	sp_human:Q9V615	- 175.00	199.02	0.0007	1338	! Q9V615 homo sapiens (human). r
	sp_human:Q96RR2	- 174.50	198.42	0.0008	1337	! Q96RR2 homo sapiens (human). m
	sp_human:Q9BZG0	- 174.50	198.41	0.0008	1338	! Q9BZG0 homo sapiens (human). r
	sp_human:Q96RG4	- 174.50	198.41	0.0008	1338	! Q96RG4 homo sapiens (human). r
	sp_human:Q96RG5	- 174.50	198.40	0.0008	1339	! Q96RG5 homo sapiens (human). r
	sp_virtebrate:Q93251	+ 174.50	197.70	0.0008	1445	! Q93251 rana catesbeiana (bull
	sp_rodent:Q62775	- 174.00	207.81	0.0007	451	! Q62775 rattus norvegicus (rat).
	sp_rodent:Q920G8	- 174.00	207.14	0.0007	485	! Q920G8 rattus norvegicus (rat).
	sp_bacteria:Q53913	- 174.00	206.38	0.0007	527	! Q53913 streptomyces coriofaciens

sp_rodent:Q63870	+ 174.00	190.53	0.0010	2944	! Q63870 mus musculus (mouse)
sp_invertebrate:Q97406	+ 172.50	195.29	0.0011	1439	! Q97406 haliotis discus (a
sp_virus:Q99307	- 171.00	197.54	0.0013	924	! Q99307 epstein-barr virus (s
sp_bacteria:Q9K3E2	+ 170.50	197.86	0.0013	835	! Q9K3E2 streptomyces coelicol
sp_virus:Q65553	- 170.50	185.35	0.0017	3247	! Q65553 bovine herpesvirus 1
seq_name: sp_bacteriap:069738					
seq_documentation_block:					
ID O69738	PRELIMINARY;	PRT;	368 AA.		
AC O69738:					
DT 01-AUG-1998 (TREMBLrel. 07, Created)					
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)					
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)					
DE PPE-FAMILY PROTEIN.					
GN RV3873 OR MTU027.08.					
OS Mycobacterium tuberculosis.					
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;					
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.					
OX NCBI_TaxID=1773;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC STRAIN=H37RV;					
RX MEDLINE=98295987; PubMed=9634230;					
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,					
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekoa F.,					
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,					
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,					
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Hamlin N., Holroyd S.,					
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,					
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,					
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;					
RT "Deciphering the biology of Mycobacterium tuberculosis from the					
RT complete genome sequence."					
RL Nature 393:537-544 (1998).					
DR EMBL; AL022120; CAA17965.1; -					
DR TuberculList; RV3873; -					
DR InterPro; IPR000030; PPE.					
DR Pfam; PF00823; PPE; 1.					
KW Complete proteome.					
SQ SEQUENCE 368 AA; 37330 MW; D78F44095F658CA2 CRC64;					
alignment_scores:					
Quality: 713.00	Length: 143				
Ratio: 5.057	Gaps: 0				
Percent Similarity: 98.601	Percent Identity: 96.503				
alignment_block:					
US-09-462-480-1 x O69738					
Align seg 1/1 to: O69738 from: 1 to: 368					
1 CTGCAGCAGTGCAGCTGTTTTCAGCCAGTGGCGGCGCCGCGCGG 50					
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242					
51 CAACCCAGCGCAGGAGCAGCCGCGCAGATGGCCTCTCGGCACCA 100					
242 YAsnProAlaAspGluGlnAlaGlnMetGlyLeuLeuGlyThrSerP 259					
101 CGCTGTGCAACCATCCGCTGGTGGTGGATCAGCCGCGCGCGCGCG 150					
259 rOleuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275					
151 GGCGCTGCTGCGCGCGAGTGCCTACCTACCTGGCGCAGGTGGTGGT 200					
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292					
201 CAGCCCGCTGATGCTCAGCTGATCGAAAGCGGTTGCCCGCTCGGTGA 250					
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309					

```
251 TGCCGGCGGCTGTTCGCCGATCGTCGGTGACGGGTGGCGCGCTCCGGTG 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 GGTCCGGGAGGATGGCGGAGGTTTCGAATCCGGCGGCTCCACGAGCC 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351 GGGTCTGTCTCGCGCGGCGGCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 oglyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspGp 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401 AGGAGGACTGGGACGACGAGGAGGACTGG 429
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
359 LuAspAspTrpAspGluGluAspAspTrp 368
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

seq_name: sp_plant:Q9ZW08

seq_documentation_block:

```
AC Q9ZW08 PRELIMINARY; PRT; 891 AA.
DT 01-MAY-1999 (TrEMBLrel: 10, Created)
DT 01-MAY-1999 (TrEMBLrel: 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel: 19, Last annotation update)
DE PUTATIVE PROLINE-RICH PROTEIN.
GN Ar2G29210.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID:3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Unayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004561; AAC95214.1; -.
DR InterPro; IPR002483; PWI.
DR InterPro; IPR002965; P-rich_extensn.
DR Pfam; PF01480; PWI; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00311; PWI; 1.
SQ SEQUENCE 891 AA; 102261 MW; C0CCB6DE05E528B CRC64;
```

alignment_scores:

```
Quality: 210.50 Length: 380
Ratio: 1.126 Gaps: 23
Percent Similarity: 49.211 Percent Identity: 31.316
```

alignment_block:

```
US-09-462-480-1/rev x Q9ZW08 ..
Align seg 1/1 to: Q9ZW08 from: 1 to: 891
1253 TCTCAGAGTGGCTCAACACTATATAACACGAGAAGGCGGAGACCGACGG 1204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 SerArgSerProIleArgArg.....HisArgArgProThrHisGluGl 312
```

```
1203 AAGTCGAACTCGCCCGGATCCCGGTGTTTCGGTATTCTACGGAACTCGGC 1154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 yArgArgGlnSerProAlaPro.....SerArgArgArg 324
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1153 GTTCCCTATCGGAACATCCAGTCACGTTGCCCTTCGGTCGAGCCATTG 1104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 rgSerProSerProProAlaArgArg.....ArgSerProSerPro 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1103 CCTCAGCGCTTCGTCGATCCGCGCGCAGGTTCCTGCAGCGCGTGTTC 1054
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 ProAlaArgArgArgSerProSerProProAlaArgArgHisArgSe 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1053 AGTCGGTAGCGGCGCTCCCATTTTGTGGACACCTGGTACGCTC 1004
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
355 rProThrProProAlaArgGln.....ArgArgSerProSerProPro 370
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1003 CGAACCGCTACCGC...CCAGGCGCTGCGAGCTTGTGTCAGGGGACTGCT 957
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 laArgArgHisArgSerProProProAlaArgArgArgArgSerProSer 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
956 TCCCTCGTCAAGGAGGGAATGAATGGACGTGACATTTCCCTGGATTGCG 907
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
387 ProProAlaArgArgArgSerProSerProProAlaArgArgArgArg 403
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
906 CTTCGCGCGGCTCGATACCGCGGAAATCCACTGCTGCTCTGTCAATTT 857
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403 gSerProSerProLeuTyArgArgAsnArgSerPro..SerProLeuTy 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
856 TTTGCTCGGTTTCITTCGTATTAGCGGTGAGAGCCCATTTGCGAGGA 807
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
419 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 431
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
806 CAGCGCTGCTGCTGCTCT...CGTCGCGCTCGAGTATTGGAGCGCGG 760
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
431 rgGlyArgSerAspSerProGlyArgSerProSerProValAlaArgLeu 447
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
759 CCTGACGAAATTCGTCGAGA.....TCTGTCGAGT 728
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
448 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 464
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 TCCGTGCTCTGCTATTGCTGCTTGGTGGAGCGCACCCG...CGCG 681
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 uProSerProProValAlaGlnArgLeuProSerProProProArgArg 481
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
680 CTGGCGCGCTCCCGCGCGCGCGCTGCGCTGCAACGAACCTG 631
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 laGlyLeuProSerProPro.....ProAlaGlnArgLeu 492
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
630 CCGTCGACTCCACTGTCGATCTGGGTTTTCAGGTGCGCGGAGA..... 586
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 ProSerProProArgArgAlaGlyLeuProSerProMetArgIleGl 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
585 .....TCCGCTCGAAATTCCTG..... 568
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
509 yGlySerHisAlaAlaAsnHisLeuGluSerProSerProSerSerLeuS 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
567 ..CCTCCTGCGCGGAGGTAGCGCATCGGCTTCATCTCTGCCATGCTGG 520
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
526 erProProGlyArg.....LysLysValLeuProSerProProValArg 540
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
519 ACTACTTTCTCTCTTTACCTTCCTCGCAAAATTTGGCAAGTCTTCGG 470
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 ArgArgArgSerLeuThrProAspGluGluArgValSerLeuSerGln... 556
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 CCGGGTGGCGGGAAGTCTGTTGTCATTAC.....GGGAGC. 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
557 ....GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 572
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
432 ..TCACCACTGCTCTCTGTCCTCCAGTCGCTCTGCTCTCTCTTCTCAG 385
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
572 etSerProVal.ArgGlyArgGlyLysSerProSerSerArg.HlsG 588
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
384 CTCTGCGGAGCGGTGCGCGCGGCGGACGACCGG.....GGCTGG 344
```

```
|||||... ||| .....
588 lnLysAlaArgSerProValArgArgSerProThrProValAsnArg 604
ID 09FVB2 PRELIMINARY; PRT; 894 AA.
AC Q9FVB2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SRM102.
GN SRM102.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lopato S., Weber K., Kalyana M., Meissner M., Langhammer U., Barta A.;
RT "AtRSP31 functions in a complex with a hSRM160-like protein.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ292982; CAC03679.1;
DR InterPro; IPR002483; PWI.
DR InterPro; IPR002965; P-rich_extensn.
DR Pfam; PF01480; PWI; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00311; PWI; 1.
SQ SEQUENCE 894 AA; 102378 MW; 705C5DC2527A6C75 CRC64;

alignment_scores:
Quality: 210.50 Length: 380
Ratio: 1.126 Gaps: 23
Percent Similarity: 49.211 Percent Identity: 31.316

alignment_block:
US-09-462-480-1/rev x Q9FVB2 ..
Align seg 1/1 to: Q9FVB2 from: 1 to: 894
1253 TCTCAGAGTCGGCTCAACAGCTATAACACGAGAAAGGCGAGACCGG 1204
|||||... ||| .....
301 SerArgSerProIleArgArg.....HisArgArgProThrHisGluG 315
1203 AAGTGCAGACTCGCCGATCCGCTGTTGGTATCTACGGAACTCGG 1154
|||||... ||| .....
315 YArgArgGlnSerProAlaPro.....SerArgArgArg 327
1153 GTTGCCCTATCGAATCCAGTGACGTTGCTTGGTGGTGAAGCCATTG 1104
|||||... ||| .....
327 IgSerProSerProAlaArgArg.....ArgSerProPro 341
1103 CCTGACCGCTCGCTGCTCGCCGACAGTTCTGACGCGCTGTTGTC 1054
|||||... ||| .....
342 ProAlaArgArgArgSerProSerProProAlaArgHisArgSe 358
1053 AGTCGGTAGCGGTGGCTCCATTTTGTGGACACCTCGGTACGCTC 1004
|||||... ||| .....
358 rProThrProProAlaArgGln.....ArgArgSerProProAla 373
```

```
1003 CGAACCGCTACCG...CCGAGCGCGCTGCGAGCTTGGTCAGGACTGCT 957
|||||... ||| .....
373 laArgArgHisArgSerProProAlaArgArgArgArgSerProSer 389
|||||... ||| .....
956 TCCCTCGTCAAGGAGGAATGAATGACGTGACATTTCCCTGGATTGCG 907
|||||... ||| .....
390 ProProAlaArgArgArgArgSerProProProAlaArgArgArgAr 406
|||||... ||| .....
906 CTTCGCGCGCTCGATACCGCGGAATTCACATGCTGCTCTGTCATGTT 857
|||||... ||| .....
406 gSerProSerProLeuTyrArgArgAsnArgSerPro..SerProLeu 422
|||||... ||| .....
856 TTTGCTCCGTTTCTTTTCGTATTAGCGGTTCAGAAGCCCATTTGCGAG 807
|||||... ||| .....
422 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 434
|||||... ||| .....
806 CAGCGCTGCTGCTGCTCT...CGTCGCGCTGCTGAGTATTGGACGCGG 760
|||||... ||| .....
434 rgGlyArgSerAspSerProGlyArgSerProSerProValAlaArgLe 450
|||||... ||| .....
759 CTGACGAATATTCGTCGAGA.....TCTGCTCGAGT 728
|||||... ||| .....
451 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 467
|||||... ||| .....
727 TCTGCTGCTGCTTATTGGCTGCTTCTTGAAGCGCACACCG...CGGC 681
|||||... ||| .....
467 uProSerProProValAlaGlnArgLeuProSerProProProArgArg 484
|||||... ||| .....
680 CTGGCGCGCTGCCCGCGCGCGCCACTGGCCCTGCAACGAACTG 631
|||||... ||| .....
484 laGlyLeuProSerProPro.....ProAlaGlnArgLeu 495
|||||... ||| .....
630 CGTCGACTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 586
|||||... ||| .....
496 ProSerProProProArgAlaGlyLeuProSerProMetArgIleG 512
|||||... ||| .....
585 .....TCCGCTCGAAATACCTT..... 568
|||||... ||| .....
512 yGlySerHisAlaAlaAsnHisLeuGluSerProSerProSerSerLeu 529
|||||... ||| .....
567 ..CCTCTCGCGCGAGGTAGCGGCATCGGCTTCATCTCTGCCATGCTGG 520
|||||... ||| .....
529 erProProGlyArg.....LysLysValLeuProProProValArg 543
|||||... ||| .....
519 ACTACTTCTCTCTTACCTTCTCTGCCCAAATGTTGCCAAGTCTTCGG 470
|||||... ||| .....
544 ArgArgArgSerLeuThrProAspGluArgValSerLeuSerGln.. 559
|||||... ||| .....
469 CCGGCTGCGCGGGAAGTCTGTTGTCATTAC.....GGGAGC. 433
|||||... ||| .....
560 ....GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 575
|||||... ||| .....
432 ..TCACAGTCGCTCTCTGCTCCAGTCGCTGCTGCTGCTGCTGCTGCTC 385
|||||... ||| .....
575 etSerProVal..ArgGlyArgGlyLysSerSerProSerSerArg.HisG 591
|||||... ||| .....
384 CTCCTGCGCGAGCGGTGCGCGCGGACGACCGG.....GGCTGG 344
|||||... ||| .....
591 lnLysAlaArgSerProValArgArgSerProThrProValAsnArg 607
|||||... ||| .....
343 TGGAGCGCGGATTGCGAACCTTGGCCCATCGCTCGCGCACCCACCGGA 294
|||||... ||| .....
608 ArgSerArgArgSerSerSerAlaSerArgSer...ProAspArgArgAr 623
|||||... ||| .....
293 GCGCGCGCACCGC.....TACCCAGCATCCGGCAACAGCGCG 256
|||||... ||| .....
623 gArgArgSerProSerSerArgSerProSerArgSerArgSerSerProp 640
|||||... ||| .....
255 CGGCA.....TCACCGAGGGGCAA 236
|||||... ||| .....
640 roValLeuHisArgSerProSerProArgGlyArg 651
```

seq_name: sp_human:Q9UQ39

seq_documentation_block:

ID Q9UQ39 PRELIMINARY; PRT; 956 AA.
AC Q9UQ39;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE RNA BINDING PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016088; BAA83714.1;
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 1 1
FT NON_TER 956 956
SQ SEQUENCE 956 AA; 106462 MW; C9FEAD07B49F5011 CRC64;

alignment_scores:

Quality: 203.50 Length: 442
Ratio: 0.988 Gaps: 27
Percent Similarity: 46.606 Percent Identity: 28.733

alignment_block:

US-09-462-480-1 x Q9UQ39

Align seg 1/1 to: Q9UQ39 from: 1 to: 956

33 GGGCGGACCGCGCGCGGCAACCCAGCGAGGGA...AGCCCGCGAGA 79
||| |||||::: :::: ||| :::: |||||:::
413 GlySerHisArgGluLeSerSerSerProThrSerLysAsnArgSerHi 429
80 TGGCGCTGCTCGGCACAGTCCGCTGTGCAACCATCCGCTGGTGGTGA 129
||| |||||::: :::: ||| :::: |||||:::
429 sGlyArgAlaLysArgAspLysSer.HisSerHisThr..... 441
130 TCAGGCCCCAGCGCGCGCGGCTGCTCGCGCGAGTCTACCTGG 179
||||| ||| |||||::: :::
442ProSerArgArgMetGlyArgSerArgSerProAlaThrAlaLy 456
180 CGCAGGTGGTGGTGTACCGCGCGCGCTGATGCTCAGCTCATCGAAA 229
||| ||| :::: ||||| ||| :::: |||||:::
456 sArgGlyArgSerArgSerArgThrProThrLysArg.GlyHisSerArg 472
230 AGCGGTGGCCCTCGGTGATGCCGCGGCTGTGCGCGATCGTCGGTG 279
||||| |||
473 SerArgSerPro..... 476
280 ACGGGTGGCGCGCTCGCGTGGGTCCGG.....GAGCGATGGG 317
::: ||| ||| :::: |||||:::
477GlnTrpArgArgSerArgSerArgAlaGlnArgTrpG 488
318 CCAGGGTTCGATTCG.....GCGGCTCCACCGACCC 349
::: ||| |||||:::
488 lyArgSerArgSerProGlnArgArgGlyArgSerArgSerProGlnArg 504
350 CGGGTCTGTCGCGCGCGCACCGCTCCGCGAGGCGTGAAGAGCGAC 399
||||| |||||::: :::: |||||:::
505 ProGlyTrpSer.....ArgSerArgAsnThrGln.ArgArgGlyA 518
400 GAGGACGACTGGGACGAAGAGGAGCTGGTGTGAGCTCCCGTAATGACAAC 449
||::: ||| :::: ||||| ||| :::: |||||:::
518 rgSerArgSerAlaArgArgGly.....ArgSerHisSer 529
450 AGACTTCCCGCGCACCGCGCGGAGACTTGCACACATTTTGGCGGAGA

seq_name: sp_human:Q9UQ40

||| ||||| ||||| ||||| |||||
530 Arg.SerProAlaThrArg...GlyArg..... 537
500 AGGTAAAGAGAGAAAGTAGTCCAGCATGGCAGAGATGAAGACCGATGCCG 549
::: |||::: :::: |||||
538SerArgSerArgThrProAla..... 544
550 CTACCCCTCGGCGAGGAGGAGGAGTAAATTCGAGCGGATCTCGGGGACCTG 599
||| ||| :::: |||||
545ArgArgGlyArgSerArgSerArgThrProAlaArgAr 557
600 AAAACCCAGATCGACACAGGTGGAGTGCAGCGCAGGTTCGTCGAGGGCCA 649
::: ||||| :::: |||||::: ||| |||||:::
557 gArgSerArgSerArgThrProThrArgArgSerArgSerArgThr. 573
650 GTGGCGCGCGCGCGGCGGAGCGCGCCAGCGCGCTGGTGGCTTCC 699
||| ||||| ||||| |||||
574ProAlaArgArgGlyArg.....Ser 580
700 AAGAACGAGCAATAAGCAGAACGAGCACTCGACGAGATCTCGACGAAT 749
::: |||::: ||||| |||||
581 ArgSerArgThrProAlaArgArgSerArgThrArgSerPro..... 595
750 ATTGCTCAGCGCGCGTCCCAATCTCGAGGCGCCAGGAGGAGCAGCAGCA 799
||| ||||| :::: |||||::: ||| |||||:::
596 ...ValArgArgArgSerArgSerArgSerProAlaArgArgSerGlyA 611
800 GGCCTGCTCTCGCAATGGCTTCTGACCGCTAATACGAAAGAAACG 849
|| |||||::: |||||
611 rgSer.....ArgSerArgThr 616
850 GAGCAAAACATGACAGAGCAGCAGTGGAAATTCGCGGCTATCGAGCGC 899
::: |||::: ||||| |||||
617 ProAlaArgArgGlyArgSerArgSerArgThrProAlaArgArgGlyAr 633
900 CGGCAAG...CGCAATCCAGGGAATGTACAGTCCATTCATTCCTCCTT 946
||| |||||::: |||||
633 gSerArgSerArgThrProAlaArg..... 641
947 GACGAGGGAAGCAGTCCCTGACCAAGCTCCAGCGCGCTGGGCGGTAG 996
||| |||||::: |||||
642ArgSerGlyArgSerArgSer 648
997 CGGTTCGAGCGGTACAGCGGTGTCCAGCAAAATGGAGCGCCACGCGTA 1046
||| |||||::: |||||
649 Arg.....ThrProAlaArgArgGlyArgSerArgSe 659
1047 CCGAGCTGAACACGCGCT..... 1065
|||::: |||||:::
659 rArgThrProArgArgGlyArgSerArgSerArgSerLeuValArgArgG 676
1066GCAGAACCTGGCGCGGAGCATCAGCAAGCCGTCAGGCAATGGC 1110
::: |||::: |||||::: |||||
676 lyArgSerHisSerArgThrProGlnArgArgGlyArgSerGlySerSer 692
1111 TTCACCGAAGCAACGTCACCTGGGATGTCGATAGGCAACCGCGAGT 1160
::: |||::: |||||
693 SerGluArgLysAsnLysSer.....ArgThrSerGlnArgArgSe 706
1161 TCGGCTAGATAAGCAAA.....ACAGGGATCGGGGAGGTTCGACC 1201
||| |||||::: |||||
706 rArgSer.AsnSerSerProGluMetLysLysSerArgLysSerSer... 721
1202 TTCGCTGGTCTCGCCCTTCTCGTGTATTACGTTTGAGCGCAGCTCTGA 1251
||| ||||| |||||
722 ...ArgArgSerArgSerLeuSerSerProArgSerLysAlaLysSer.. 736
1252 GAGGTGTCATGGCGG 1267
||| ||||| |||
737 .ArgLeuSerLeuArg 741

```

seq_documentation_block:
ID Q9UQ40 PRELIMINARY; PRT; 1262 AA.
AC Q9UQ40;
DT 01-MAY-2000 (TReMBLrel_13, Created)
DT 01-MAY-2000 (TReMBLrel_13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel_19, Last annotation update)
DE RNA BINDING PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ontaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RR EMBL: AB016087; BAA83713.1;
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS; PRO1217; PRICHEXTENS.
FT NON_TER 1
FT NON_TER 1262
FT SEQUENCE 1262 AA; 138003 MW; 75E76BD73FA14295 CRC64;

alignment_scores:
Quality: 203.50 Length: 442
Ratio: 0.988 Gaps: 27
Percent Similarity: 46.606 Percent Identity: 28.733

alignment_block:
US-09-462-480-1 x Q9UQ40 ..
Align seg 1/1 to: Q9UQ40 from: 1 to: 1262

33 GGGCGGCACCGGGCGGCGCAACCCAGCCGACGAGGA...AGCCGCGCAGA 79
||| ||||| ||||| ||||| ||||| ||||| |||||
238 GlySerHisArgGluIleSerSerProThrSerLysAsnArgSerHI 254
||| ||||| ||||| ||||| ||||| ||||| |||||
80 TGGCCCTGCTCGGCACACGACGCTCTGCAACCATCCGCTGGCTGGTGG 129
||| ||||| ||||| ||||| ||||| ||||| |||||
254 sGlyArgAlaLysArgAspLysSer.HisSerHisThr..... 266
||| ||||| ||||| ||||| ||||| ||||| |||||
130 TCAGGCCCGCAGCGGGCGGCGGCTGCTGCGCGCGAGTGCCTACCTGG 179
||||| ||||| ||||| ||||| ||||| ||||| |||||
267 .....ProSerArgArgMetGlyArgSerArgSerProAlaThrAlaLys 281
||| ||||| ||||| ||||| ||||| ||||| |||||
180 CGCAGGTGGTGTGTGACCGCGCACGCGCTGATGCTCAGCTGTAGTAA 229
||| ||||| ||||| ||||| ||||| ||||| |||||
281 sArgGlyArgSerArgSerArgSerArgProThrLysArg.GlyHisSerArg 297
||| ||||| ||||| ||||| ||||| ||||| |||||
230 ACCCGGTGGTGGCCCTCGGTGATGCGCGGCGCTGTTGCCGGATCGTCGGTG 279
||||| ||||| ||||| ||||| ||||| ||||| |||||
298 SerArgSerPro..... 301
||| ||||| ||||| ||||| ||||| ||||| |||||
280 ACCGGTGGCGCGCGCTCGGTGGTGGTGGG.....GAGCGATGGG 317
||| ||||| ||||| ||||| ||||| ||||| |||||
302 .....GlnTrpArgArgSerArgSerArgSerAlaGlnArgTrpG 313
||| ||||| ||||| ||||| ||||| ||||| |||||
318 CCAGGGTTCGAATCCG.....GCGGCTCCACCCACCC 349
||| ||||| ||||| ||||| ||||| ||||| |||||
313 LysArgSerArgSerProGlnArgArgGlyArgSerArgSerProGlnArg 329
||| ||||| ||||| ||||| ||||| ||||| |||||
350 CGGGTGTGGTGGCGCGCGGCGGCTGACCGCTCGCAGGAGCGGTGAAGAACGAC 399
||||| ||||| ||||| ||||| ||||| ||||| |||||
330 ProGlyTrpSer.....ArgSerArgAsnThrGln.ArgArgGlyA 343
||| ||||| ||||| ||||| ||||| ||||| |||||
400 GAGCAGCTGGGACGAAGAGGACGACTGGTGAGCTCCCGTAATGACAAAC 449
||||| ||||| ||||| ||||| ||||| ||||| |||||
343 rgSerArgSerAlaArgArgGly.....ArgSerHisSer 354
||| ||||| ||||| ||||| ||||| ||||| |||||
450 AGACTTCGCGCACCGCGGCGCGGAGACTTGCACATATTTGGCGCAGA 499
||| ||||| ||||| ||||| ||||| ||||| |||||
355 Arg.SerProAlaThrArg...GlyArg..... 362
||| ||||| ||||| ||||| ||||| ||||| |||||

```



```
241 CCCTCGGTGATCCGGCGGCTGTTCCCGGATCGTCGGTGAAGGTGGCGC 290
|||||
544 ProGln.....ArgGluArgSerGlySerValAspG1 557
291 CGCTCCGGTGGTCCGGGAGGATGGCCAGGGTTCGCAATCCGGCGCT 340
|||||
557 nLysThrValAlaArgThrProLeuGlnArgSerArgSerGlySers 574
341 CCACAGCCCGGCTCGTCCGCGCGGACCGCTCGCGCAGGAGCGTGA 390
|||
574 erGlnGluLeuAsp...ValLysProSerAlaSerProGlnGluArgSer 589
391 GAAGCAGCAGGAGCAGCTGGGACGAGAGCAGCTGGTGAGCTCCCGT 440
|||||
590 GluSer.....AspSerSerProAspSerLysAlaLysTh 601
441 AATGACAACACATCCCGCGCCACCGCGGCGGAGAGCTCCCAACATT. 489
: |||
601 rArgThrProLeuArgGlnArgSerArgSerGlySerSerProGluVal 618
490 .....TTGCGAGGAAGTAAAGAGAGAAAGTAGTCCAGCA 525
|||||
618 spSerLysSerArgLeuSerProArgArgSerArgSerGlySerSerPro 634
526 TGCAGAGATGAACCCGATCCGCTACCCCTCGGCGAGGAGGAGTAA 575
|||
635 GluValLysAspLysProArgAlaAlaProArgAlaGlnSerGlySer 651
576 TTCAGCGGATCTCCGGGACCTGAACACCCAGATCG...ACCAGTGG 622
|||||
651 pSerSerProGluProLysAlaProAlaProArgAlaLeuProArgS 668
623 GTCAGGCGAGGTCGTTGTCAGGCGCAGTGGCGGCGCGCGGAGCG 672
|||||
668 erArg.....SerGlySerSerLysGlyArg 677
673 CCGCCCGCGCGGTGGTGGCTTCCAGAAAG.....CAGCCAA 713
|||
678 GlyProSerProGluGlySerSerSerThrGluSerSerProGluHis 694
714 AAGCAGAGCAGGACATCGACGATCTCGACGAATATTCGTACAGCGG 763
|||||
694 oProLysSerArgThrAlaArgArgGlySerArgSerSerProGlu 711
764 CGTCCAATACGCA...GGGCGCAGGAGGACGACGACGCGCTGCTCT 810
|||||
711 ysThrLysSerArgThrProProArgArgArgSerSerSerSerPro 727
811 .....CGCAAAATGGGCTTC 824
728 GluLeuThrArgLysAlaArgLeuSerArgArgSerArgSerAlaSer 744
825 TGACCCGCTATAGCAAAAGAACGAGGAGAAACATGACAGACGACAG 874
|||||
744 rSerProGluThrArgSerArgThrProProArgHisArgArg..Ser 760
875 TGGAAAT.....TCGCGGTATCGAGCGCGC 900
|||||
760 oSerValSerSerProGluProAlaGluLysSerArgSerArgArg 777
901 GGCAAGCGCAATCCAGGAAATGTCAGCTCATTCCTCCTCCTTCAGC 950
|||||
777 rArgSerAlaSerSerProArgThrLys.....ThrThrSerArg 790
951 AGGGAGAGCAGTCCCTGACCAAGC..... 974
791 ArgGlyArgSerProSerProLysProArgGlyLeuGlnArgSerArg 807
975 ....TCCAGCGGCGCTGGGCGGTTCGGGTCGAGCGGTACAGGGTGT 1020
|||||
807 rArgSerArgGluLysThrArgThrThrArgArgArgAspArgSerG 824
```

```
1021 CCAGCAAAATGGAGCCACCGCTACCGAGCTGAACAACCGCTGCAGA 1070
|||||
824 lySerSerGlnSerThrSerArg.ArgArg.....GlnArgSerArgSe 838
1071 ACCTGGCGCGGACGATCAGCGAAGCCGTCAGGCAATGGCTTCGACCG 1120
: |||
838 rArgSerArgValThrArgArgArgArgGlySerGlyTyrHisSerA 855
1121 GGCAAA.....CGTCACTGGGATGTCGCATAGGCAACCGCGAGTTCGC 1164
|||||
855 rgSerProAlaArgGlnGluSerSerArgThrSerSerArgArgArg 871
1165 GTAGATAGCGAAACA..... 1180
872 Gly.ArgSerArgThrProProThrSerArgLysArgSerArgSerArg 888
1181 .....CGGGATCGGCGGAGT...TCGACC 1201
888 hrSerProAlaProTrpLysArgSerArgSerArgAlaSerProAlaThr 904
1202 TTCGTCGGTCTCGC 1216
905 HisArgArgSerArg 909
seq_name: sp_human:Q9H5F9
seq_documentation_block:
ID Q9H5F9 PRELIMINARY; PRT; 704 AA.
AC Q9H5F9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA: FLJ23471 FIS, CLONE HS111969.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027124; BAB15667.1;
DR InterPro: IPR002965; P_rich_extensn.
DR PRINTS: PRO1217; PRICHEXTENS.
SQ SEQUENCE 704 AA; 75768 MW; 858F94EEA2C1F8C6 CRC64;
```

alignment_scores:

Quality: 193.00 Length: 504
Ratio: 0.881 Gaps: 30
Percent Similarity: 43.452 Percent Identity: 25.992

alignment_block:

US-09-462-480-1/rev x Q9H5F9 ..

Align seg 1/1 to: Q9H5F9 from: 1 to: 704

```
1217 GCGAGACCGACGAGGAGTGAAC.....TCGCCGATCCGCTGT 1177
|||||
92 GlyAsnSerProAlaArgAlaSerValProAlaAlaProAsnProAlaAl 108
1176 TCGTATTCTACGGAATCGG...CGTTGCCCTATGCGAACATCCGAGT 1130
|||||
108 aThrSerAlaThrSerValHisValArgSerProAlaArgProSerGlu 125
1129 GAGTTCCCTTCGTCGAAGCCATTGCTCGCGGCTTCGC..... 1089
|||||
125 erArg...LeuAlaProThrProThrGluGlyLysValArgProArgVal 140
```

```
1088 TGATCGTCGCGCCCA.....GGTCTCAGCGGCTTGTTCAGCTCG 1048
      :::::|||||
141 ThrAsnSerProMetGlyTrpSerAlaAlaProCysThrAla... 156
      :::::|||||
1047 GTAGCCGTGGCGTCCCATTTTGTGTGACACCCCTGGTACGCTCCGGAACC 998
      :::::|||||
157 .....AlaAlaAlaSerHisProAlaValProSerA 168
      :::::|||||
997 GCTACCGCCCGCAGCGCGTGGTGGTCTGAGGACTGTCTCCCTCCCTCGT 948
      :::::|||||
168 laProAspProArgPro.....AlaThrProGln 177
      :::::|||||
947 CAAGGAGGAATGAATGACGCTGACATTTCCCTGGATTGCGCTTGCCTGCG 898
      :::::|||||
178 GlyGlyGly.....AlaProArgValAlaAlaProG1 188
      :::::|||||
897 GCCTCGATACCGCGGAATTCACCTGCTGCTCTCATGTTTTTTCCTCCG 848
      :::::|||||
188 nThrThrLeuSerSerSerThrSerAlaAlaThr.ValAspProPro 204
      :::::|||||
847 TTTCTTTTCTGATTAGCGGGTCAGAAGCCCATTTTCCGAGGACA..... 805
      :::::|||||
205 AlaIrpThrProSerAlaSer.ArgThrGlnGlnAlaArgAsnLysPheP 221
      :::::|||||
804 ..GGCCCTGCTGCTGCTCTCGTCCGCCCTCGAGTATTGGACGCGCGCT 757
      :::::|||||
221 heGlnThrSerAlaValProGlyThrSerLeuSerGlyArgGlyPro 237
      :::::|||||
756 .....GACCAATATTCTCGAGATCTCGTCG 731
      :::::|||||
238 ThrProSerLeuValLeuSerLysAspSerSerLysGluGlnAlaArgAs 254
      :::::|||||
730 AGTTCTCTGCTCTGCTATTGGCTGCTTCTTGGAGGCGCACCCGCGGC 681
      :::::|||||
254 nPheLeuLysGlnAlaLeuSerAlaLeuGluAlaGlyAlaProAlap 271
      :::::|||||
680 CTGGCGCGCGCTCCCGCGCGCGCGCGCC.....ACTGGCGCTGCA 640
      :::::|||||
271 roGlyArgProSerProAlaThrAlaAlaValProSerSerGlnProLys 287
      :::::|||||
639 ACGAACCTGCGG.....TCGACTCC 620
      :::::|||||
288 ThrGluAlaProGlnAlaSerProLeuAlaLysProLeuGlnSerSerSe 304
      :::::|||||
619 ACCTGGTGGATCTGGTTTTCAGTCCGCGGAGA..... 586
      :::::|||||
304 rProArgValLeuGlyLeuProSerArgMetGluProProAlaProLeuS 321
      :::::|||||
585 .....TCCGCTCGAAATACCTGCCTCCT..... 562
      :::::|||||
321 erThrSerSerThrSerGlnAlaSerAlaLeuProProAlaGlyArgArg 337
      :::::|||||
561 .....GCCGAGGTACGCGCATCGGCTTCATCTCTGCCATCTGGAC 518
      :::::|||||
338 AsnLeuAla.GluserSerGlyValGly.....ArgValGlyA 350
      :::::|||||
517 TACTTTCT.....CTCTTT 504
      :::::|||||
350 laGlySerArgProLysProGluAlaProMetAlaLysGlyLysSerThr 366
      :::::|||||
503 ACCTTCCTCGCCCAAAATTTGGCAAGTCTT..... 474
      :::::|||||
367 ThrLeuThrGlnAspMetSerThrSerLeuGlnGluGlyGlnGluAspG1 383
      :::::|||||
473 ..CCGCGCGCGGTGCGCGGAAGTCTG.....TTCTCATACGGGAGCTC 431
      :::::|||||
383 yProAla..GlyTrpArgAlaAsnLeuLysProValAspArgArgSerPro 399
      :::::|||||
430 ACCAGTCTGCTCTTCGTCGTCAGTCTGCTCTGCTCTCTT..... 389
      :::::|||||
400 AlaGluArgThrLeuLysProlysGluProArgAlaLeuAlaGluProAr 416
      :::::|||||
388 .....CACGCTCTGCGGAGCG..... 371
```

```
      :::::|||||
416 gAlaGlyGluAlaProArgLysValSerGlySerPheAlaGlySerValH 433
      :::::|||||
370 .....GTCCGCGCGGACAGACCCGCGGCTGGTGGAGCGCGCG 332
      :::::|||||
433 isIleThrLeuThrProValArgProAspArg.....ThrProArg 446
      :::::|||||
331 ATTGCGAACCTGCGCCATCGTCCCGGACCCACCGGAGCGCGCCACCC 282
      :::::|||||
447 ProAlaSerProGlyProSerLeuPro.....AlaArgSerPr 459
      :::::|||||
281 GTCACCGACGATCCGGCAACAGCCCGCGGCATCACCGAGGGGCAACCGG 232
      :::::|||||
459 oSerProProArg..... 464
      :::::|||||
231 CTTTTCGATCATCGATGACATCATCGCGGCGTCAACGACCCACCTG 182
      :::::|||||
464 ..... 464
      :::::|||||
181 CGCCAGGTAGCTAGCTCCGCGGACAGCGCCGCGCGCTGGGCGCT 132
      :::::|||||
465 ArgArgLeuAlaValProAlaSerLeuAspValCysAspAsnTrpLeuAr 481
      :::::|||||
131 GATCCACGACCGACGCG...GATGGTTGCACAGCGGACTGGTGGCGAGCAG 85
      :::::|||||
481 gProGluProProGlyGlnGluAlaArgValGlnSerTrpLysGluGluG 498
      :::::|||||
84 GCCCATCTGCGCGGCTTCTCTGCTGCGTGGTGGTGGCGCGCGGTTGCCG 35
      :::::|||||
498 luLysLysProHisLeuGlnGlyArgProGly.....Arg 509
      :::::|||||
34 CCA.....CCTGGCTGAACAAGCAGCTC 12
      :::::|||||
510 ProLeuSerProAlaAsnValProAlaLeuProGlyGluThrValThrSe 526
      :::::|||||
11 ACCT 8
      :::::|||||
526 rPro 527

seq_name: sp_bacteria:Q9XDH2
seq_documentation_block:
ID Q9XDH2 PRELIMINARY; PRT; 763 AA.
AC Q9XDH2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DSC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
RC MEDLINE=2090472; PubMed=10627046;
RA Espitia C., Lacleite J.P., Mondragon-Palomino M., Anador A.,
RA Campuzano J., Martens A., Singh M., Cicero R., Zhang Y., Moreno C.;
RT "The PE-PGRS glycine-rich proteins of Mycobacterium tuberculosis: a
RL Microbiology 145:3487-3495(1999).
DR EMBL; AF071081; AAD41594.1;
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistil_extensin.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01218; PSTLEXTENSIN.
SQ SEQUENCE 763 AA; 75035 MW; 39168EC45A5916F8 CRC64;
```

alignment_scores:
Quality: 193.00
Length: 435

Ratio: 1.078 Gaps: 21
Percent Similarity: 41.149 Percent Identity: 27.126

alignment_block:
US-09-462-480-1/rev x Q9XDH2 ..

Align seg 1/1 to: Q9XDH2 from: 1 to: 763

```
1124 TGCCTTCGTCGACGAGCCATTCCTGACCGGCTTCGCTGATGCTCGCGCC 1075
|||
|||
281 CysProLeuProProLeuProAsn.....AsnHisProPr 293
1074 AGGTTCTGCACGCGGTTCCTGACGTCGGTAGCGTGGCTGCCATTTTGG 1025
|::|
|||
293 oAlaProAlaProValProGlyValProLeuAlaProLeuProA 310
1024 CTGGACACCTGGTACGCTCCGACCGCTACCGCC 987
|::|
|||
310 snSerHisProProAlaProProSerAlaProValProGlyValProLeu 326
986 .....AGGCGCTGCGAGCTTGGTCCAGGACTGC 958
|||
|||
327 AlaProLeuProIleSerGlyArgProValSerValTrpLysGlySerPh 343
957 TTCC..... 954
::|
|||
343 eThrThrLeuSerThrPheCysCysArgValCysSerGlyGluValLeuA 360
953 .....CTCGTCAAGGAGGGAATGATGACGCTGACAT 921
360 laGlyAlaLeuAsnProSerArgProSerArgSerProLeuThrThr 376
920 TTCCTGATTCGCTTCGCGCGCTCGATACCGCGAAATTCACATGC 871
|||
|||
377 ThrProAlaLeuProAlaProIleProProLeuProProLeuProProL 393
870 TGCTCTGTCATGTTTTCGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 821
|||
|||
393 euProIleAsnThrAlaValProIleProProLeuProProVal.Th 409
820 CCCATTTCGAGGACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 771
409 iAlaLeuAlaProProLeuProProLeuAlaPro...LeuProIleSerP 425
770 TTGGACGCGCGCTGACGAAATTCCTGCGAGATCTCGTTCGAGTTCCTGCT 721
|||
|||
425 roGlyValProPro.....AlaProProIleProPro 435
720 TCTGCTTATTCGCTGCTT.....CTTGGAGCGCACACCGCGG..... 682
436 GlyLysProTrpThrThrProProLeuAlaProAlaProProGluProLy 452
682 ..... 682
452 sThrValProValLeuProProGlyProSerCysProSerGluLysP 469
681 .....CTGGGCGCGCTCCCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTG 640
469 roAsnProProAlaProProGluProProGluProLysSerSerProAla 485
639 ACGAACTCGCGTGCATCCACCTGGTGGATCTCGGTTTCAGTTCGCGCG 590
486 .....LeuProProAlaProProAlaProSerMetProSerAlaValAr 500
589 GAGATCCGCTCGAAATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
500 gValProProSerProIleProProAlaProProAlaProProAlaProAr 517
539 CTTCATCTCTGCCATGCTGGACTACTTCTCTCTCTCTCTCTCTCTCTCT 490
517 laSer.....MetProAlaLeu 522
489 AATGTTGGCAAGTCTTCGCGCGCGCGGAGTGTCTGTTCTCATTA 440
```

```
523 ProProAlaProProSerProProAlaThrArg...LeuCysProProLe 538
439 CGGAGCTCACCAGT.....CGTCTCTTCTGTCCTCCA 409
538 uProProSerProProAlaProAsnSerProProAlaProProAlaProP 555
408 GTCGCTCCTGCTGCTTCTTCCAGCTCCTGCGGAGCGGTGCGCGCGCA 359
555 roThrProLysLeuLeuSerAlaAsnProProCysProProValPro 571
358 CCACACCGCGGCTGGTGGAGCGCGGATTTGGGAACCTTGGC..... 317
572 ProAlaPro.....AsnArgProProAlaProProAlaProProAl 585
316 ...CCATCGCTCCGCGCACCGGCGCGCGCGCGCGCGCGCGCGCGCG 271
585 aProProGluLeuProAlaProProAsP.....ProProThrProPro 600
270 TCCGGCACACCGCGCGCATCACGAGGCGCGCGCGCGCGCGCGCGCG 221
600 alAlaAsnSerProProAlaProProAla.....ProPro 611
220 GCTGAGACATCAGCGCGTGGGGTCAACGACCGCGCGCGCGCGCGCG 171
612 AlaProProSerAla...LeuProPheValAsnProProAlaProProThr 627
170 GACTCCGCG...CGCAGCAGCGCGCG...CCGCGCGTGGCGCGCTGATCC 127
628 ProAlaAlaProLysSerArgProAlaLeuProAlaAlaProProAlaPr 644
126 ACCAGCCAGCGGATGTTTCGACGCGGACTGTCGCGCGCGCGCGCGCT 77
644 oProAlaProProValArgAlaThrThrProProProAlaProProAlaP 661
76 GCGCGGCTTCTCTGCTGCGTGGTGGTGGCG.....CGCGCGTGGCGCG 33
661 roProAlaProAsnSerMetAlaLeuProProAlaProProAspProPro 677
```

seq_name: sp_rodent:070495

seq_documentation_block:
ID 070495 PRELIMINARY; PRT; 897 AA.
AC 070495;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PLENTY-OF-PROLINES-101.
GN SRRM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Vaysiere B.M., Camonis J.H.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF062655; AAC17422.1;
DR MGD; MGI:1858303; Srrm1.
DR InterPro; IPR002483; PWI.
DR Pfam; PF01480; PWI; 1.
DR SMART; SM00311; PWI; 1.
SQ SEQUENCE 897 AA; 101166 MW; 62160BEB6772BD10 CRC64;

alignment_scores:
Quality: 192.00 Length: 563
Ratio: 0.869 Gaps: 27
Percent Similarity: 39.254 Percent Identity: 25.933

alignment_block:
US-09-462-480-1/rev x 070495 ..

Align seg 1/1 to: 070495 from: 1 to: 897

```
1271 TCGCCGCGCATGACAACCTCTCAGAGTGGCTCAACAGCTATAACACGAG 1222
282 SerArgSerArgThrArgSerArgSerProSerHisThrArgProArgArg 298
1221 AAAGGGGAGACGACGAGGAGTGCAGAACTCGCCCGCATCCCGTGTTCGCT 1172
298 gArgHisArgSerArgSerArgSerTyrSerProArgArg.....Argp 313
1171 ATTTACGCGAACHTCGGGTTCCTATGCGAACAATCCAGTAGCTGTC 1122
313 roSerProArgArgArgProSerProArgArgThrPro..... 326
1121 CTTCGGTGAAGCATTCCTCGACCGCTTCGTCGATCGCCGCGCCAGG 1072
327 ProArgArgMetProProProProArgHisArgSerArgSerProGl 343
1071 T..... 1067
343 yArgArgArgArgSerSerAlaSerLeuSerGlySerSerSers 360
1066 CAGCGCGTTGTTACAGTCGGTAGCGGTGGCTGCCATTTTCGTGGACAC 1017
360 erSerSerArgSerArgSerPro..... 368
1016 CCTGGTACGCTCCGAACCGCTACCGC.....CCAGCGCGCTCGGAGC 973
369 ProLysLysProProLysArgThrSerSerProProArgLysThrArgAr 385
972 TTGTCAGGACTGCTCCCTCCCTCAAGGAGGAATGAATGACGTGAC 923
385 gLeuSerProSerAlaSerProProArgArgArgHisArgProSer.... 400
922 ATTTCCCTGATTCGGCTTCGCGCGGCTCGATAC.....CCGCGA 882
401 ..SerProAlaThrProProProLysThrArgHisSerProThrProGln 416
881 AATTCCACTGCTGCTCTCATGTTTTTGCTCCGTTCT...TTTCGTAT 835
417 GlnSerAsnArgThrArgLysSerArgValSerValSerProGlyArgTh 433
834 TAGCGGTTCAGAACCCATTTCCAGGAGCA.....G 804
433 rSerGlyLysValThrLysHisLysGlyThrGluLysArgGluSerProS 450
803 CGCCTGCTGCTCCTCTGT..... 784
450 erProAlaProLysProArgLysValGluLeuSerGluSerGluGluAsp 466
784 ..... 784
467 LysGlySerLysMetAlaAlaAspSerValGlnGlnArgArgGlnTy 483
784 ..... 784
483 rArgArgGlnAsnGlnGlnSerSerSerAspSerGlySerSerSerThrS 500
783 .....CGGCC.....TCAGATTGGA 766
500 erGluAspGluArgProLysArgSerHisValLysAsnGlyGluValGly 516
765 CGCCGG..... 760
517 ArgArgArgArgLeuSerProSerArgSerAlaSerProSerProArgLy 533
760 ..... 760
533 sArgGlnLysGluThrSerProArgMetGlnMetGlyLysArgTrpGlnS 550
759 ..CCTGACGAATATTCGTGAGATCTCGTAGTTCCTGCTTCGCTTAT 712
!!! !!!!!!! !!!!! !!!!!
```

```
550 erProValThrLysSerSerArgArgArgArgSerProSerProProPro 566
711 TGGCTGCTTCTTGAAGC.....GCACCACCGCGCTCGCGGCC 671
567 AlaArgArgArgArgSerProSerProAlaProProProProProPr 583
670 GTCCCCCGCGCGCGCCACTGCGCCCTGCA.....ACGAACC 633
583 oProProProArgArgArgSerProThrProProProArgArgArgT 600
632 TCCCGTCCACTCCACCTGGTCGATCTGGTTTTCAGTCCGCGGAGATCC 583
600 hrProSerProProProArgArgSerProSerProArgArgTyrSer 616
582 .....GCTCGAAA 575
617 ProProLleGlnArgArgTyrSerProSerProProProLysArgArgTh 633
574 TTACTGCTCTCTGCCGAGGTAGCGCATCGGTCTTCATCTCTGCCAT 525
633 rAlaSerProProProProLysArgArgAlaSerProSerProPro. 649
524 GCTGGACTACTTCTCTCTTTTACCTTCTCGCCCAAAATGTTGGCAAGTCT 475
650 .....ProLysArgArgValSerHisSer 657
474 TCCGCGCGGCTGCGCGGAAGTCTGTTGTCATTACGGGAGCTACCACT 425
658 ProProProLysGlnArgSerProThrValThrLysArgArgSerPro.S 674
424 CGTCTCTCTGCTCCAGTCGTCCTCGTCTTCTTTCACGCTCTCGCGC 375
674 erLeuSerSerLysHisArgLysGlySerSerProGlyArgSer.ThrAr 690
374 AGCGGTGCGCGCGCACCGACCGCGGTGTTGGAGCCGCGGATTGCGA 325
690 gGluAlaArgSer...ProGlnProAsn.....LysArgHisSerProS 704
324 ACCCTGCGCCCATCGCTCCCGAC.....CCACCGAGCGCGC 287
704 erProArgProArgAlaProGlnThrSerSerProProProValArgArg 720
286 .....CACCGTCACCGACGATCCGGCAACAGCGCGCGGATCACCGAG 243
721 GlyAlaSerAlaSerProGlnGlyArgGlnSerPro.....SerProSe 735
242 GGGCAACCGCGCTTTCGATCAGCTGAGACA..... 212
735 rThrArgProIleArgArgValSerArgThrProGluProLysLysIleL 752
211 ..TCAGCGCGTGGGTCAACAGCACCCACCTCGCCGAGTAGCGACTCCG 164
752 ysLysAlaAlaSerProSerProGlnSerValArgArgValSerSer 768
163 CGCGCAGCAGG.....CCCGCGCGCGCTGGGGCTGATCCACACAG 121
769 ArgSerValSerGlySerProGluProAla...AlaLysLysProProAl 784
120 CAGCGGATGTTTCGACAGCGGACTGGTCCGAGCAGGAGGCCATCTCGCGG 71
784 a.....ProSerProValGlnSerG 792
70 CTTCCTCGTCCGCTGGGTTCGCGCGCGCGGTGCCG 36
792 lnSerProSerThrAsnTrpSerProAlaValPro 803
seq_name: sp_human:Q9UQ36
seq_documentation_block:
ID Q9UQ36 PRELIMINARY; PRT; 1275 AA.
AC Q9UQ36;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
```


seq_name: sp_human:Q9Y2J6

seq_documentation_block:

```

ID Q9Y2J6 PRELIMINARY; PRT; 772 AA.
AC Q9Y2J6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA0992 PROTEIN (FRAGMENT).
GN KIAA0992.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
DR EMBL; AB023209; BAA76836.1; -.
DR HSSP; P56276; ITLK.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002965; P-rich_extensn.
DR Pfam; PF00047; 19; 3.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00408; IGC2; 3.
KW Immunoglobulin domain.
FT NON_TER 1
SQ SEQUENCE 772 AA; 83643 MW; EE5EE2BC79C7492B CRC64;

```

alignment_scores:

Quality	Length
Ratio: 1.381	312
Percent Similarity: 42.949	Gaps: 15
	Percent Identity: 28.526

alignment_block:

US-09-462-480-1/rev x Q9Y2J6 ..

Align seg 1/1 to: Q9Y2J6 from: 1 to: 772

```

822 AGCCCATTTGCGAGG.....ACAGCGCTGCTGCTGCTCC 788
||||| :|||
3 SerProGlyGlyArgGlyIleLysProAspThrCysProAlaProGlypr 19
787 TCGTCGGCCC.....TCGAGTA 771
||||| |||
19 oArgSerProLeuGlnLeuProLeuAlaProAspAlaGluSerGlySers 36
770 TTGGACGGCGGCTGAC..... 754
||||| |||
36 erGlyArgArgProGlyCysProArgaspProLeuLysLeuGlnGlnLeu 52
753 .....GAATATTCGTGAGATCTCGTCGAGTTCTGTC 722
||||| :|||
53 GlnAsnGlnIleArgLeuGluGlnGluAlaGlyAlaArgGlnProPropr 69
721 TTCGTCTATTGGCTGCTTCTTGAAGCGCACCGCGGCTGCGGCGG 672
::: ||||| |||
69 oAlaPro.....ArgSerAlaProProSerProProPhep 81
671 CGTCCCGCGCGCGCGCCACTGGCCCTGCAACGAACCTGCGCTGACT 622
|| ||||| |||
81 roProProAlaPheProGluLeuAlaAlaCysThrProProAlaSer 97
621 CCACCTGCTGCTGGGTTTTCAGTGTGCGCGGAGATCCGCTCGA..... 577
||| ||||| |||

```

```

98 ProGluProMetSerAlaLeuAlaSerArgSerAlaProAlaMetGlnSe 114
576 .....AATTACTGCTCTCCGCCGAGGGTAGCGGCATCGGT 540
||||| ||| :|
114 rSerGlySerPheAsnTyrAlaArgProLysGln..... 125
539 CTTTCATCTCTGCCATGCTGGACTACTTTCTCTCTTTTACCTTTCCTCGCCAA 490
::: ||||| :|||
126 .....PheileAlaAlaGln 130
489 AATGTTGGCAAGTCTTCGGGCC..GGTGGCGGGGAAGTCTGTCTCATTT 441
||||| ||| :|||
131 AsnGlyProAlaSerGlyHisGlyThrProAlaSerProSerSe 147
440 ACGGAGCTCACAGTCTCTCTCTGTCGCCAGTCTCTGCTGCTCTTC 391
::: ||||| ||| :|||
147 rSerSerLeuProSer....ProMetSerProThr....ProArgGlnPheG 162
390 TTCAGCTCTCT.....GGCGAGCGGTGCCGGCGCGACAGAC 353
::: ||||| ||| :|||
162 LyArgAlaProValProPheAlaGlnProPheGlyAlaGluProGlu 178
352 CCGGCTGGTGGAGCGCGGATTCGGAACCTGCGCCATCGCTCCCGGA 303
||| ||| :|||
179 AlaProTrpGlySer.....SerSerProSerProProProPr 192
302 CCCA.....CGGAGCGCGCCACCGCTCACCGAGATCCGGC 265
||| ||| ||| :|||
192 oProValPheSerProThrAlaAlaPheProValProaspValPheP 209
264 AACAGCGCGCGCATCACCGAGGGGCAACCGGCTTTTCGATCAGCTGAG 215
||||| |||
209 roLeuProProProPro..... 215
214 ACATCAGCGCGGTGCGGTCAACGACCCACCTGCCCGCAGTAGCGACTCC 165
||| :||| ||| :|||
216 .....ProLeu..ProSerProGlyGlnAlaSer 224
164 GCGCGCAGCAGCGCGCGCGCTGGG...CCTGATCCACCGAGCAG 118
||| ||||| :|||
225 HisCysSerSerProAlaThrArgPheGlyHisGlyGlnThrProAlaAl 241
117 CGGATGTTGCACAGCGGACTGGTCCGAGCAGGCGCCATCTCGCGGCTT 68
: ||| ||||| :|||
241 a.....PheLeuSerAlaLeuLeuProSerGlnProProProAlaAlav 256
67 CCGTGGTGGGTGGGTGGCGCGCGGTGGCGGCC 33
::: ||||| ||| |||
256 alAsnAlaLeuGlyLeuProLysGlyValThrPro 267

```

Mon Jul 22 08:29:58 2002

us-09-462-480-1.rspt

Page 16